GenCore version 5.1.3

OM protein - protein search, using sw model

Run on: February 20, 2003, 10:46:10 ; Search time 12 Seconds
(without alignments)
48.067 Million cell updates/sec

Title: US-09-912-414-9
Perfect score: 31
Sequence: 1 WXXWXF 6

Scoring table: BLOSUM62DX
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_73:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	31	100.0	21	2	PH1731	Ig heavy chain V r
2	31	100.0	49	2	I48681	185 kDa glycophosp
3	31	100.0	59	2	E82844	hypothetical prote
4	31	100.0	61	2	H91196	hypothetical prote
5	31	100.0	65	2	S23164	light-harvesting p
6	31	100.0	71	2	F84168	hypothetical prote
7	31	100.0	73	2	AH2788	hypothetical prote
8	31	100.0	76	2	T14724	hypothetical prote
9	31	100.0	78	2	S08449	hypothetical prote
10	31	100.0	79	2	A64750	hypothetical prote
11	31	100.0	83	2	A97568	hypothetical prote
12	31	100.0	86	2	T42185	conserved hypothet
13	31	100.0	87	2	H86753	prophage pi2 prote
14	31	100.0	92	2	S37509	Ig kappa chain V r
15	31	100.0	92	2	JQ1627	hypothetical 10.7K
16	31	100.0	93	2	G87098	hypothetical prote
17	31	100.0	97	2	E84147	hypothetical prote
18	31	100.0	99	2	C84022	hypothetical prote
19	31	100.0	101	2	S73518	probable lipoprote
20	31	100.0	103	2	F82266	conserved hypothet
21	31	100.0	105	2	JE0384	NADH2 dehydrogenas
22	31	100.0	106	2	I64171	hypothetical prote
23	31	100.0	106	2	C82715	conserved hypothet
24	31	100.0	106	2	G72012	transport permease
25	31	100.0	108	2	AG1164	hypothetical prote
26	31	100.0	108	2	S28241	NADH2 dehydrogenas
27	31	100.0	108	2	AG2505	hypothetical prote
28	31	100.0	109	2	F83057	hypothetical prote
29	31	100.0	109	2	AI2029	hypothetical prote
30	31	100.0	110	2	AD0127	conserved hypothet
31	31	100.0	110	2	AG1523	hypothetical prote
32	31	100.0	116	2	B28966	Ig heavy chain pre
33	31	100.0	116	2	G82398	conserved hypothet
34	31	100.0	117	2	A95381	hypothetical prote
35	31	100.0	118	2	S24527	Ig heavy chain V r
36	31	100.0	119	2	S24522	Ig heavy chain V r
37	31	100.0	119	2	S24492	Ig heavy chain V r
38	31	100.0	119	2	S24513	Ig heavy chain V r
39	31	100.0	119	2	S24493	Ig heavy chain V r
40	31	100.0	119	2	S24491	Ig heavy chain V r
41	31	100.0	119	2	S24525	Ig heavy chain V r
42	31	100.0	119	2	S24520	Ig heavy chain V r
43	31	100.0	119	2	S24518	Ig heavy chain V r
44	31	100.0	119	2	S24511	Ig heavy chain V r
45	31	100.0	119	2	S24494	Ig heavy chain V r

ALIGNMENTS

RESULT 1
PH1731
Ig heavy chain V region (clone GCC-11) - mouse (fragment)
C;Species: Mus musculus (house mouse)
C;Date: 24-Feb-1994 #sequence_revision 24-Feb-1994 #text_change 17-Mar-1999
C;Accession: PH1731
R;McHeyzer-Williams, M.G., McLean, M.J., Lalor, P.A., Nossal, G.J.V.
J. Exp. Med. 178, 295-307, 1993
A;Title: Antigen-driven B cell differentiation in vivo.
A;Reference number: PH1675; MUID:93301607; PMID:8315385
A;Accession: PH1731
A;Molecule type: mRNA
A;Residues: 1-21 <MCH>
A;Experimental source: B cell
A;Note: the authors translated the codon ACA for residue 13 as Ala
C;Superfamily: immunoglobulin V region; immunoglobulin homology
C;Keywords: heterotetramer; immunoglobulin

Query Match 100.0%; Score 31; DB 2; Length 21;
Best Local Similarity 50.0%; Pred. No. 1.3e+02;
Matches 3; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 WXXWXF 6
|..|.|
Db 16 WTSWSF 21

RESULT 2
I48681
185 kDa glycophosphoprotein - mouse (fragment)
C;Species: Mus musculus (house mouse)
C;Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 03-Jun-2002
C;Accession: I48681
R;White, M.R.; Hung, M.C.
Oncogene 7, 677-683, 1992
A;Title: Cloning and characterization of the mouse neu promoter.
A;Reference number: I48681; MUID:92228496; PMID:1348855
A;Accession: I48681
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-49 <RES>
A;Cross-references: EMBL:X66236, NID:g395186, PIDN:CAA46966.1, PID:g395187
C;Genetics:
A;Gene: neu
C;Keywords: ATP; phosphoprotein

Query Match 100.0%; Score 31; DB 2; Length 49;
Best Local Similarity 50.0%; Pred. No. 2.7e+02;
Matches 3; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 WXXWXF 6
|..|.|

```
Db      6 WCRWGF 11


RESULT 3
E82844
hypothetical protein XF0107 [imported] - Xylella fastidiosa (strain 9a5c)
C;Species: Xylella fastidiosa
C;Date: 16-Aug-2000 #sequence_revision 16-Aug-2000 #text_change 16 Aug 2000
C;Accession: E82844
R;anonymous; The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequ
Nature 406, 151-157, 2000
A;Title: The genome sequence of the plant pathogen Xylella fastidiosa.
A;Reference number: A82815; MUID:20847717; PMID:10910347
A;Note: For a complete list of authors see reference number A82820 below
A;Accession: E82844
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-50 <SIM>
A;Cross-references: GP:AE003866, GB:AE003849, NID:g9104304, PIDN:AAF83339.1, GSPDB:GN0
A;Experimental source: strain 9a5c
F;Campos, A.J.G.; Reinach, F.C., Arruda, P., Abreu, F.A., Acencio, M., Alvarenga, R., A
Briones, M.R.S., Bruno, M.R.P., Camargo, A.A., Camargo, L.E.A., Carraro, D.M., Carrer, H
as Neto, E., Dorea, C., Ferro, M.I., Furlaneti, A.P., Ferreira, A.C.S.
submitted to GenBank, June 2000
A;Authors: Ferreira, V.C.A.; Ferro, J.A., Fraga, J.S., Franco, S.C., Franco, M.C., Frolm
J.D.; Junqueira, M.L.; Kemper, E.L., Kitajima, J.P., Krieger, J.E., Kuramae, E.E., Laign
chado, M.A.; Madeira, A.M.B.N., Madeira, H.M.F., Marino, C.L., Marques, M.V., Martins, E
A;Authors: Martins, E.M.F., Matsukuma, A.Y., Menck, C.F.M., Miracca, E.C., Miyaki, C.Y.,
, F.C., Nunes, L.R., Oliveira, M.A., de Oliveira, M.C., de Oliveira, R.C., Palmieri, D.A
Rodrigues, V., Rosa, A.J., de M., de Rosa Jr., V.E., de Sa, R.G., Santelli, R.V., Sawasak
A;Authors: da Silva, A.C.R., da Silva, E.F., da Silva, A.M., Silva Jr., W.A., da Silveir
M., Tsukahi, M.H., Valadao, H., Van Sluys, M.A., Verjovski-Almeida, S., Vettore, A.L., Z
A;Reference number: A69328
A;Contents: annotation
C;Genetics:
A;Gene: XF0107

  Query Match             100.0%;  Score 31;  DB 2;  Length 50;
  Best Local Similarity   100.0%;  Pred. No. 3.1e+02;
  Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 WXXWXF 6
           |:|:|
Db      54 WRCWLF 59


RESULT 4
H91196
hypothetical protein ECs4544 [imported] - Escherichia coli (strain O157:H7, substrain EC
C;Species: Escherichia coli
C;Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
C;Accession: H91196
R;Hayashi, T., Makino, K., Ohnishi, M., Kurokawa, K., Ishii, K., Yokoyama, K., Han, C.G
gashwara, N., Yasunaga, T., Kuhara, S., Shiba, T., Hattori, M., Shinagawa, H.
DNA Res. 8, 11-22, 2001
A;Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and geno
A;Reference number: A39023; MUID:21156231; PMID:11258796
A;Accession: H91196
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-61 <HAY>
A;Cross-references: GB:BA000007; PIDN:BAB37967.1; PID:g13364019; GSPDB:GN00164
A;Experimental source: strain O157:H7, substrain RIMD 0509952
C;Genetics:
A;Gene: ECs4544

  Query Match             100.0%;  Score 31;  DB 2;  Length 61;
  Best Local Similarity   100.0%;  Pred. No. 3.4e+02;
  Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 WXXWXF 6
           |:|:|
Db       3 WDRWFF 8


RESULT 5
S23164
light-harvesting protein alpha chain - Ectothiorhodospira halochloris
N;Alternate names: antenna pigment protein alpha chain
C;Species: Ectothiorhodospira halochloris
C;Date: 14-Mar-1997 #sequence_revision 14-Mar-1997 #text_change 11-Aug-1998
C;Accession: S23164
R;Wagner-Huber, F., Prunishole, P.A., Biesig, T., Frank, G., Suter, F., Zuber, H.
Eur. J. Biochem. 205, 917-925, 1992
A;Title: The primary structure of the antenna polypeptides of Ectothiorhodospira halochl
A;Reference number: S23164; MUID:92243236; PMID:1577699
A;Accession: S23164
A;Molecule type: protein
A;Residues: 1-65 <WAG>
C;Superfamily: light-harvesting protein alpha chain
C;Keywords: antenna complex; bacteriochlorophyll; blocked amino end; light harvesting pol
F;1/Modified site: N-formylmethionine #status experimental

  Query Match             100.0%;  Score 31;  DB 2;  Length 65;
  Best Local Similarity   100.0%;  Pred. No. 3.5e+02;
  Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 WXXWXF 6
           |:|:|
Db       2 WKLWKF 7


RESULT 6
F84168
hypothetical protein Vng1072h [imported] - Halobacterium sp. NRC-1
C;Species: Halobacterium sp. NRC-1
C;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C;Accession: F84168
R;Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B., Pan, M., Shukla, H.D., Lasky, S
, Leithauser, B., Keller, K., Cruz, R., Danson, M.J., Hough, D.W., Maddocks, D.G., Jablon
Jung, P.H., Alam, M., Freitas, T.
Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000
A;Authors: Ng, W., Daniels, C.J., Tolmie, R.I., Cruz, A.I., Ebhardt, H., Lowe, T.M., Lie
A;Title: Genome sequence of Halobacterium species NRC-1.
A;Reference number: AF4160; MUID:20504183; PMID:11016950
A;Accession: F84168
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-71 <STO>
A;Cross-references: GB:AE004437; NID:g15789781; PIDN:AAG19714.1; GSPDB:GN00138
C;Genetics:
A;Gene: VNG0670H

  Query Match             100.0%;  Score 31;  DB 2;  Length 71;
  Best Local Similarity   100.0%;  Pred. No. 3.8e+02;
  Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 WXXWXF 6
           |:|:|
Db      49 WWLWGF 54


RESULT 7
AH2788
hypothetical protein Atu1707 [imported] - Agrobacterium tumefaciens (strain C58, Dupont)
C;Species: Agrobacterium tumefaciens
C;Date: 11-Jun-2002 #sequence_revision 11-Jun-2002 #text_change 11 Jun 2002
C;Accession: AH2788
R;Wood, D.W., Setubal, J.C., Kaul, R., Monks, D., Chen, L., Wood, G.E., Chen, Y., Woo, L
etage, G., Miller, W., Brunt, P., Duenthmer, D., Matyassik, T., Levy, R., Li, M., McClella
, Karg, K., Romero, E., Chang, C.
Science 294, 2317-2323, 2001
A;Authors: Yuan, H., Tao, Y., Biddle, P., Jiang, M., Fraser, W., Purry, M., Gordon-Kamm, B
ster, E.W.
A;Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A;Reference number: AB2077; PMID:11743193
```

A;Accession: AH2788
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-73 <KUR>
A;Cross-references: GB:AE008698; PIDN:AAL42726.1; PID:g17740165; GSPDB:GN00186
A;Experimental source: strain C58 (Dupont)
C;Genetics:
A;Gene: Atu1727
A;Map position: circular chromosome

Query Match 100.0%; Score 31; DB 2; Length 73;
Best Local Similarity 50.0%; Pred. No. 3.9e+02;
Matches 3; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 WXXWXF 6
|::|:|
Db 13 WNLWIF 18

RESULT 8
T14724
hypothetical protein Y1091 - Yersinia pestis plasmid pMT1
C;Species: Yersinia pestis
C;Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 24-Sep-1999
C;Accession: T14724; T15006
R;Hu, P.; Elliott, J.; Mcready, P.; Skowronski, E.; Garnes, J.; Kobayashi, A.; Carrano,
submitted to the EMBL Data Library, March 1998
A;Description: Structural organization of virulence determinants in three Yersinia pesti
A;Reference number: Z18168
A;Accession: T14724
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-76 <HUP>
A;Cross-references: EMBL:AF053947, NID:g2996296, PID:g2996363, PIDN:AAC13243.1
R;Lindler, L.E.; Plano, G.V.; Burland, V.; Mayhew, G.F.; Blattner, F.R.
Infect. Immun. 66, 5731-5742, 1998
A;Title: Complete DNA sequence and detailed analysis of the Yersinia pestis KIM5 plasmid
A;Reference number: Z18268; MUID:99043898; PMID:9826348
A;Accession: T15006
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-76 <LIN>
A;Cross-references: EMBL:AF074611, NID:g3883083, PID:g3883089, PIDN:AAC82749.1
C;Genetics:
A;Gene: Y1091
A;Genome: plasmid pMT1

Query Match 100.0%; Score 31; DB 2; Length 76;
Best Local Similarity 50.0%; Pred. No. 4e+02;
Matches 3; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 WXXWXF 6
|::|:|
Db 50 WRTWNF 55

RESULT 9
S08449
hypothetical protein 9 - spiroplasma virus 1
C;Species: spiroplasma virus 1, SpV1
C;Date: 30-Jun-1991 #sequence_revision 30-Jun-1991 #text_change 07-Dec-1999
C;Accession: S08449
R;Renaudin, J.; Aullo, P.; Vignault, J.C.; Bove, J.M.
Nucleic Acids Res. 18, 1293, 1990
A;Title: Complete nucleotide sequence of the genome of Spiroplasma citri virus SpV1-R8A2
A;Reference number: S08447; MUID:90206799; PMID:2320423
A;Accession: S08449
A;Status: translation not shown
A;Molecule type: DNA
A;Residues: 1-78 <REN>
A;Cross-references: EMBL:X51344; NID:g61393; PIDN:CAA35713.1; PID:g998898
C;Genetics:
A;Genetic code: SGC3

Query Match 100.0%; Score 31; DB 2; Length 78;
Best Local Similarity 50.0%; Pred. No. 4.1e+02;
Matches 3; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 WXXWXF 6
|::|:|
Db 42 WLIWLF 47

RESULT 10
A64750
hypothetical protein b0249 - Escherichia coli (strain K-12)
C;Species: Escherichia coli
C;Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 01-Mar 2002
C;Accession: A64750
R;Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Col
.A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A;Title: The complete genome sequence of Escherichia coli K-12.
A;Reference number: A64720; MUID:97426617; PMID:9278503
A;Accession: A64750
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-79 <BLAT>
A;Cross-references: GB:AE000133; GB:U00036; NID:g1787699; PIDN:AAC73255.1; PID:g1786113;
A;Experimental source: strain K-12, substrain MG1655

Query Match 100.0%; Score 31; DB 2; Length 79;
Best Local Similarity 50.0%; Pred. No. 4.1e+02;
Matches 3; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 WXXWXF 6
|::|:|
Db 50 WRAWNF 55

RESULT 11
A97568
hypothetical protein AGR_C_3173 [imported] - Agrobacterium tumefaciens (strain C58, Cere
C;Species: Agrobacterium tumefaciens
C;Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 11-Jan-2002
C;Accession: A97568
R;Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman,
A.; Liu, P.; Wallam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.;
Science 294, 2323-2328, 2001
A;Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum
A;Reference number: A97359; PMID:11743194
A;Accession: A97568
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-83 <KUR>
A;Cross-references: GB:AE007869; PIDN:AAK87498.1; PID:g15156835; GSPDB:GN00169
C;Genetics:
A;Gene: AGR_C_3173
A;Map position: circular chromosome

Query Match 100.0%; Score 31; DB 2; Length 83;
Best Local Similarity 50.0%; Pred. No. 4.3e+02;
Matches 3; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 WXXWXF 6
|::|:|
Db 23 WNLWIF 28

RESULT 12
T42185
conserved hypothetical protein L7085 - Escherichia coli plasmid pO157
C;Species: Escherichia coli
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C;Accession: T42185
R;Burland, V.; Shao, Y.; Perna, N.T.; Plunkett, G.; Sofia, H.J.; Blattner, F.R.

```
Nucleic Acids Res. 26, 4196-4204, 1998
A;Title: The complete DNA sequence and analysis of the large virulence plasmid of Esche
A;Reference number: S27065; MUID:98391744; PMID:9722640
A;Accession: T42185
A;Status: preliminary; translated fr  GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-86 <PUR>
A;Cross-references: EMBL:AF004613; PIDN:AAC70153.1
A;Experimental sou  e: strain EDL933, serotype O157:H7
C;Genetics:
A;Genome: plasmid pO157
A;Note: L7085

  Query Match             100.0%;  Score 31;  DB 2;  Length 86;
  Best Local Similarity    50.0%;  Pred. No. 4.5e+02;
  Matches    3;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 WXXWXF 6
           |::|:|
Db      62 WRTWNF 67


RESULT 13
H86753
prophage pi2 protein 25 [imported] - Lactococcus lactis subsp. lactis (strain IL1403)
C;Species: Lactococcus lactis subsp. lactis
C;Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 03-Aug-2001
C;Accession: H86753
R;Bolotin, A.; Windyer, P.; Maiget, F.; Tilli S.; C.; Malarme, P.; Ehrli
Genome Res. 11, 731-753, 2001
A;Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis su
A;Reference number: A86625; MUID:21235186; PMID:11337471
A;Accession: H86753
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-87 <STO>
A;Cross-references: GB:AE006176; NID:g12723963; PIDN:AAK05150.1; GSPDB:GN00146
A;Experimental source: strain IL1403
C;Genetics:
A;Gene: pi225

  Query Match             100.0%;  Score 31;  DB 2;  Length 87;
  Best Local Similarity    50.0%;  Pred. No. 4.5e+02;
  Matches    3;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 WXXWXF 6
           |::|:|
Db      17 WSAWAF 22


RESULT 14
S37509
Ig kappa chain V region (V-kappa 3) - human (fragment)
C;Species: Homo sapiens (man)
C;Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 21-Jul-1999
C;Accession: S37509
R;Klein, U.; Kueppers, R.; Rajewsky, K.
submitted to the EMBL Data Library, September 1993
A;Description: Human IgM(+)IgD(+) cells, the major B cell subset in the peripheral blood
A;Reference number: S37501
A;Accession: S37509
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-92 <KLE>
A;Cross-references: EMBL:Z29003; NID:g405659; PIDN:CAA82356.1; PID:g405660
C;Superfamily: immunoglobulin V region; immunoglobulin homology
C;Keywords: heterotetramer; immunoglobulin

  Query Match             100.0%;  Score 31;  DB 2;  Length 92;
  Best Local Similarity    50.0%;  Pred. No. 4.7e+02;
  Matches    3;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 WXXWXF 6
           |::|:|
Db      77 WRTWTF 82


RESULT 15
JQ1627
hypothetical 17.7K protein - turkey rhinotracheitis virus
C;Species: turkey rhinotracheitis virus
C;Date: 17-Apr-1998 #sequence_revision 17-Apr-1998 #text_change 13-Oct-1999
C;Accession: JQ1627
R;Ling, R.; Easton, A.J.; Pringle, C.R.
J. Gen. Virol. 73, 1709-1715, 1992
A;Title: Sequence analysis of the 22K, SH and G genes of turkey rhinotracheitis virus and
A;Reference number: S_4.1; MUID:92368037; PMID:1629937
A;Accession: JQ1627
A;Molecule type: mRNA
A;Residues: 1-92 <LIN>
A;Cross-references: GB:S44191; NID:g1164.1; PIDN:AAB23948.1; PID:g291706

  Query Match             100.0%;  Score 31;  DB 2;  Length 92;
  Best Local Similarity    50.0%;  Pred. No. 4.7e+02;
  Matches    3;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 WXXWXF 6
           |::|:|
Db      65 WAYWFF 70


Search completed: February 21, 2003, 11:11:01
Job time : 13 secs
```

GenCore version 5.1.3

OM protein - protein search, using sw model

Run on: February 20, 2003, 19:41:55 ; Search time 6.66667 Seconds
(without alignments)
37.329 Million cell updates/sec

Title: US-09-912-414-9
Perfect score: 31
Sequence: 1 WXXWXF 6

Scoring table: BLOSUM62DX
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution

SUMMARIES

Result No	Score	% Query Match	Length	DB	ID	Description
1	31	100.0	65	1	LHA2_ECTHL	P80103 ectothiorho
2	31	100.0	78	1	VG9_SPV1P	P15900 spiroplasma
3	31	100.0	79	1	YKFF_ECOLI	P75677 escherichia
4	31	100.0	84	1	ACHA_CROPS	P54248 crocidura r
5	31	100.0	84	1	ACHA_FELCA	P54250 felis silve
6	31	100.0	90	1	YBJI_ECOLI	P58095 escherichia
7	31	100.0	101	1	YG50_MYCPN	P75147 mycoplasma
8	31	100.0	105	1	NIGM_HUMAN	Q95178 homo sapien
9	31	100.0	106	1	YQCC_HAEIN	Q57152 haemophilus
10	31	100.0	108	1	NIGM_BOVIN	Q02374 bos taurus
11	31	100.0	116	1	HV05_CARAU	P19181 carassius a
12	31	100.0	128	1	MERT_STAAU	P08656 staphylococ
13	31	100.0	129	1	KV3H_HUMAN	P04207 homo sapien
14	31	100.0	129	1	YHW2_YEAST	P38857 saccharomyc
15	31	100.0	138	1	YA91_MYCPN	P75602 mycoplasma
16	31	100.0	139	1	YE13_MYCPN	Q9exd6 mycoplasma
17	31	100.0	144	1	HV26_MOUSE	P01795 mus musculu
18	31	100.0	145	1	PL32_AERPE	Q9yf92 aeropyrum p
19	31	100.0	146	1	YG56_YEAST	P53311 saccharomyc
20	31	100.0	147	1	Y565_METJA	Q57985 methanococc
21	31	100.0	150	1	HOPD_ECOLI	P08932 escherichia
22	31	100.0	157	1	YE63_MYCPN	P75320 mycoplasma
23	31	100.0	164	1	LSPA_ECOLI	P00804 escherichia
24	31	100.0	165	1	DSBB_VIBAL	Q56578 vibrio algi
25	31	100.0	165	1	LSPA_ENTAE	P13514 enterobacte
26	31	100.0	165	1	YPH1_PHOSH	Q53229 rhodobacter
27	31	100.0	166	1	Y18K_SSV1	P20209 sulfolobus
28	31	100.0	170	1	LSPA_PSEFL	P17942 pseudomonas
29	31	100.0	171	1	Y428_MYCGE	P47667 mycoplasma
30	31	100.0	172	1	YB03_MYCPN	P75566 mycoplasma
31	31	100.0	176	1	VPI_BPP2	P26701 bacteriopha
32	31	100.0	177	1	TN18_HUMAN	Q9ung2 homo sapien
33	31	100.0	182	1	VG37_BPMU	Q9t1v8 bacteriopha
34	31	100.0	196	1	YJP8_YEAST	P46989 saccharomyc
35	31	100.0	200	1	YBG1_MYCTU	P71689 mycobacteri
36	31	100.0	206	1	YMJF_ECOLI	P37626 escherichia
37	31	100.0	208	1	YHHN_ECOLI	P37616 escherichia
38	31	100.0	210	1	SSH5_YEAST	Q03446 saccharomyc
39	31	100.0	213	1	IF4E_XENLA	P48597 xenopus lae
40	31	100.0	215	1	IF4E_APLCA	Q77210 aplysia cal
41	31	100.0	217	1	IF4E_HUMAN	P06730 homo sapien
42	31	100.0	217	1	IF4E_MOUSE	P09415 mus musculu
43	31	100.0	217	1	IF4E_RABIT	P29328 oryctolagus
44	31	100.0	219	1	FNGP_METJA	Q57768 methanococc
45	31	100.0	220	1	RP28_BPSP1	P03048 bacteriopha

ALIGNMENTS

RESULT 1
LHA2_ECTHL
ID LHA2_ECTHL STANDARD; PRT; 65 AA.
AC P80103;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 01-FEB-1994 (Rel. 28, Last annotation update)
DE Light-harvesting protein B800/820/1020, alpha-2 chain (FHS-alpha 2,
DE (Antenna pigment protein, alpha-2 chain).
OS Ectothiorhodospira halochloris.
OC Bacteria; Proteobacteria; Gamma subdivision; Ectothiorhodospiraceae;
OC Halorhodospira.
OX NCBI_TaxID=1052;
RN [1]
RP SEQUENCE.
RC STRAIN=DSM 1059;
RX MEDLINE=92249336; PubMed=1577009;
RA Wagner-Huber R., Brunisholz R.A., Bissig I., Frank G., Suter F.,
RA Zuber H.;
RT "The primary structure of the antenna polypeptides of
RT Ectothiorhodospira halochloris and Ectothiorhodospira halophila. Four
RT core-type antenna polypeptides in E. halochloris and E. halophila.";
RL Eur. J. Biochem. 205:917-925(1992).
CC -!- FUNCTION: ANTENNA COMPLEXES ARE LIGHT-HARVESTING SYSTEMS, WHICH
CC TRANSFER THE EXCITATION ENERGY TO THE REACTION CENTERS.
CC -!- SUBUNIT: THE CORE COMPLEX IS FORMED BY DIFFERENT ALPHA AND BETA
CC CHAINS, BINDING BACTERIOCHLOROPHYLL MOLECULES, AND ARRANGED
CC MOST PROBABLY IN TETRAMERIC STRUCTURES DISPOSED AROUND THE
CC REACTION CENTER. THE NON PIGMENTED GAMMA CHAINS MAY CONSTITUTE
CC ADDITIONAL COMPONENTS.
CC -!- SUBCELLULAR LOCATION: Type II membrane protein. Inner membrane.
DR PIR; S23164; S23164.
DR InterPro; IPR002361; AntennaA.
DR InterPro; IPR000066; Antenna_a/b.
DR Pfam; PF00556; LHC; 1.
DR PROSITE; PS00968; ANTENNA_COMP_ALPHA; 1.
KW Antenna complex; Light harvesting polypeptide; Transmembrane;
KW Magnesium; Bacteriochlorophyll; Inner membrane.
FT DOMAIN 1 13 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 14 34 POTENTIAL.
FT DOMAIN 35 65 PERIPLASMIC (POTENTIAL).
FT METAL 29 29 AXIAL LIGAND TO THE BACTERIOCHLOROPHYLL
FT MAGNESIUM (POTENTIAL).
SQ SEQUENCE 65 AA; 7688 MW; 55A4C396748E3D9A CRC64;

Query Match 100.0%; Score 31; DB 1; Length 65;
Best Local Similarity 50.0%; Pred. No. 1.3e+02;
Matches 3; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 WXXWXF 6
|..|.|
Db 2 WKLWKF 7

RESULT 2
VG9_SPV1P

```
ID   VG9_SPV1P      STANDARD;      PRT;    78 AA.
AC   P15900;
DT   01-APR-1990 (Rel. 14, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   01-FEB-1996 (Rel. 33, Last annotation update)
DE   Gene 9 protein.
GN   9.
OS   Spiroplasma virus SpV1 R8A2 B.
OC   Viruses, ssDNA viruses, Inoviridae, Plectrovirus.
OX   NCBI_TaxID=10854;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90206799; PubMed=2320423;
RA   Renaudin J., Aullo P., Vignault J.C., Bove J.M.;
RT   "Complete nucleotide sequence of the gene 9 of Spiroplasma virus
RT   virus SpV1-R8A2 B.";
RL   Nucleic Acids Res. 18:1293-1293(1990).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X51344; CAA35727.1; -.
SQ   SEQUENCE   78 AA;  8761 MW;  BF8F81CFBBBBBBBB 78CF4;

  Query Match             100.0%;  Score 31;  DB 1;  Length 78;
  Best Local Similarity    50.0%;  Pred. No. 1.6e+02;
  Matches    3;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 WXXWXF 6
             |::|:|
Db      42 WLIWLF 47


RESULT 3
YKFF_ECOLI
ID   YKFF_ECOLI     STANDARD;      PRT;    79 AA.
AC   P75677; Q93CD8;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15 JUL 1998 (Rel. 36, Last sequence update)
DT   16-OCT 2001 (Rel. 41, Last annotation update)
DE   Hypothetical protein ykfF
GN   YKFF OR B0249.
OS   Escherichia coli
OC   Bacteria; Proteobacteria; gamma subdivision, Enterobacteriaceae;
OC   Escherichia.
OX   NCBI_TaxID=562;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12 / MG1655;
RX   MEDLINE 97426017; PubMed=9278503;
RA   Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA   Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA   Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA   Mau B., Shao Y.;
RT   "The complete genome sequence of Escherichia coli K-12.";
RL   Science 277:1453-1474(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12 / W3110;
RA   Takemoto K., Mori H., Murayama N., Hatada E., Yano M., Itoh T.,
RA   Yamamoto Y., Inokuchi H., Miki T., Hatada E., Fukuda R., Ishihara S.,
RA   Mizuno T., Makino K., Nakata A., Yura T., Campel G., Miyashiro K.,
RT   "Systematic sequencing of the Escherichia coli genome: analysis of the
RT   4.0 - 6.0 min (189,987 - 291,416bp) region.",
RL   Submitted (FEB 1996) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: STRONG, TO E.COLI YPJI.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE000153, AAC73502.1; -.
DR   EMBL; D83536, BAA77018.1; ALT_INIT.
DR   EcoGene, EG14282, ykfF.
KW   Hypothetical protein, Complete proteome.
SQ   SEQUENCE   79 AA;  9014 MW;  6145D747C21A234A CRC64;

  Query Match             100.0%;  Score 31;  DB 1;  Length 79;
  Best Local Similarity    50.0%;  Pred. No. 1.6e+02;
  Matches    3;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 WXXWXF 6
             |::|:|
Db      50 WRAWNF 55


RESULT 4
ACHA_CRORS
ID   ACHA_CRORS     STANDARD;      PRT;    84 AA.
AC   P54248;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01 OCT 1996 (Rel. 34, Last sequence update)
DT   15 JUN 2002 (Rel. 41, Last annotation update)
DE   Acetylcholine receptor protein, alpha chain (Fragment).
GN   CHRNA1.
OS   Crocidura russula (White toothed shrew).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Insectivora; Soricidae; Crocidurinae; Crocidura.
OX   NCBI_TaxID=36802;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Muscle;
RX   MEDLINE 95345091; PubMed 7619817;
RA   Barchan D., Ovadia M., Kochva E., Fuchs S.;
RT   "The binding site of the nicotinic acetylcholine receptor in animal
RT   species resistant to alpha-bungarotoxin.";
RL   Biochemistry 34:9172-9176(1995).
CC   -!- FUNCTION: AFTER BINDING ACETYLCHOLINE, THE AChR EXPERIENCES AN
CC       EXTENSIVE CHANGE IN CONFORMATION THAT AFFECTS ALL SUBUNITS AND
CC       LEADS TO OPENING OF AN ION-CONDUCTING CHANNEL ACROSS THE PLASMA
CC       MEMBRANE.
CC   -!- SUBUNIT: PENTAMER OF TWO ALPHA CHAINS, AND ONE EACH OF THE BETA,
CC       DELTA, AND GAMMA (IN IMMATURE MUSCLE) OR EPSILON (IN MATURE
CC       MUSCLE) CHAINS.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein (By similarity).
CC   -!- SIMILARITY: BELONGS TO THE LIGAND-GATED IONIC CHANNEL FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U17006; AAA89113.1; -.
DR   InterPro; IPR001175; Neur_channel.
DR   Pfam; PF02931; Neur_chan_LBD; 1.
DR   PROSITE; PS00236; NEUROTR_ION_CHANNEL; 1.
KW   Receptor, Postsynaptic membrane, Ionic channel, Glycoprotein;
KW   Transmembrane.
FT   NON_TER       1      1
FT   DISULFID      7     21       BY SIMILARITY.
FT   DISULFID     71     72       ASSOCIATED WITH RECEPTOR ACTIVATION (BY
FT                                SIMILARITY).
FT   CARBOHYD     23     23       N-LINKED (GLCNAC...) (POTENTIAL).
FT   NON_TER      84     84
```

```
SQ   SEQUENCE   84 AA;  9833 MW;  232072FBC967C86B CRC64;

  Query Match             100.0%;  Score 31;  DB 1;  Length 84;
  Best Local Similarity    50.0%;  Pred. No. 1.7e+02;
  Matches    3;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 WXXWXF 6
           |::|:|
Db      63 WKHWVF 68


RESULT 5
ACHA_FELCA
ID   ACHA_FELCA     STANDARD;      PRT;    84 AA.
AC   P54250;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Acetylcholine receptor protein, alpha chain (Fragment).
GN   CHRNA1.
OS   Felis silvestris catus (Cat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Carnivora; Fissipedia; Felidae; Felis.
OX   NCBI_TaxID=9685;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Muscle;
RX   MEDLINE=95345091; PubMed=7619817;
RA   Barchan D., Ovadia M., Kochva E., Fuchs S.;
RT   "The binding site of the nicotinic acetylcholine receptor in animal
RT   species resistant to alpha-bungarotoxin.";
RL   Biochemistry 34:9172-9176(1995).
CC   -!- FUNCTION: AFTER BINDING ACETYLCHOLINE, THE ACHR RESPONDS BY AN
CC       EXTENSIVE CHANGE IN CONFORMATION THAT AFFECTS ALL SUBUNITS AND
CC       LEADS TO OPENING OF AN ION-CONDUCTING CHANNEL ACROSS THE PLASMA
CC       MEMBRANE.
CC   -!- SUBUNIT: PENTAMER OF TWO ALPHA CHAINS, AND ONE EACH OF THE BETA,
CC       DELTA, AND GAMMA (IN IMMATURE MUSCLE) OR EPSILON (IN MATURE
CC       MUSCLE) CHAINS.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein (By similarity).
CC   -!- SIMILARITY: BELONGS TO THE LIGAND-GATED IONIC CHANNEL FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U17007; AAA89114.1; -.
DR   InterPro; IPR001175; Neur_channel.
DR   Pfam; PF02931; Neur_chan_LBD; 1.
DR   PROSITE; PS00236; NEUROTR_ION_CHANNEL; 1.
KW   Receptor; Postsynaptic membrane; Ionic channel; Glycoprotein;
KW   Transmembrane.
FT   NON_TER       1      1
FT   DISULFID      7     21       BY SIMILARITY.
FT   DISULFID     71     72       ASSOCIATED WITH RECEPTOR ACTIVATION (BY
FT                                SIMILARITY).
FT   CARBOHYD     20     20       N-LINKED (GLCNAC...) (POTENTIAL)
FT   NON_TER      84     84
SQ   SEQUENCE   84 AA;  9823 MW;  5B6E9A10BA1BBD7E CRC64;

  Query Match             100.0%;  Score 31;  DB 1;  Length 84;
  Best Local Similarity    50.0%;  Pred. No. 1.7e+02;
  Matches    3;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 WXXWXF 6
           |::|:|
Db      63 WKHWVF 68
```

```
RESULT 6
YPJI_ECOLI
ID   YPJI_ECOLI     STANDARD;      PRT;    90 AA.
AC   P58095;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Hypothetical protein ypjI.
GN   YPJI OR B2641.1
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
OX   NCBI_TaxID=562;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12 / MG1655;
RX   MEDLINE=97426617; PubMed=9278503;
RA   Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA   Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA   Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA   Mau B., Shao Y.;
RT   "The complete genome sequence of Escherichia coli K-12.";
RL   Science 277:1453-1474(1997).
RN   [2]
RP   IDENTIFICATION, AND CONCEPTUAL TRANSLATION.
RA   Rudd K.E.;
RL   Unpublished observations (JAN-1999).
CC   -!- SIMILARITY: STRONG, TO E.COLI YPFF.
CC       !: CAUTION: THIS IS A CONCEPTUAL TRANSLATION. A FRAMESHIFT HAD TO BE
CC       CORRECTED IN POSITION 69 AND TWO IN FRAME STOP CODONS HAD TO BE
CC       SUPPRESSED IN POSITIONS 8 AND 11.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE000349; -; NOT_ANNOTATED_CDS.
DR   EcoGene; EG13309; ypjI.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   90 AA;  10070 MW;  E1C3E2F6C2E8B697 CRC64;

  Query Match             100.0%;  Score 31;  DB 1;  Length 90;
  Best Local Similarity    50.0%;  Pred. No. 1.8e+02;
  Matches    3;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 WXXWXF 6
           |::|:|
Db      61 WRAWNF 66


RESULT 7
YG50_MYCPN
ID   YG50_MYCPN     STANDARD;      PRT;   101 AA.
AC   P75147;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hypothetical lipoprotein MPN650 precursor (E09_orf101).
GN   MPN650 OR MP192
OS   Mycoplasma pneumoniae.
OC   Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX   NCBI_TaxID=2104;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 29342 / M129;
RX   MEDLINE=97105885; PubMed=8948633;
RA   Himmelreich R., Hilbert H., Plagens H., Pirkl E., Li B.-C.,
RA   Herrmann R.;
RT   "Complete sequence analysis of the genome of the bacterium Mycoplasma
```

```
RT   pneumoniae.";
RL   Nucleic Acids Res. 21:4117-4123(1993).
CC   -!- SUBCELLULAR LOCATION: Attached to the membrane by a lipid anchor
CC       (Potential).
CC   -!- SIMILARITY: BELONGS TO THE M3439 / M3441 FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaborative
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is not in any way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE000020; AAB35840.1; -.
DR   InterPro; IPR001595; Lipoprotein_3.
DR   Pfam; PF00938; Lipoprotein_3; 1.
DR   ProDom; PD003076; Lipop: Lip_3; 1.
DR   PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.
KW   Hypothetical protein; Lipoprotein; Membrane; Signal;
KW   Complete proteome.
FT   SIGNAL       1     19       POTENTIAL
FT   CHAIN       20    101       HYPOTHETICAL LIPOPROTEIN MJ0439.
FT   LIPID       20     20       N-ACYL DIGLYCERIDE (C POTENTIAL)
SQ   SEQUENCE   101 AA;  11183 MW;  C71F0B9EF197573D CRC64;

  Query Match             100.0%;  Score 31;  DB 1;  Length 101;
  Best Local Similarity   100.0%;  Pred. No. 1.2e+03;
  Matches    3;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 WXXWXF 6
           |::|:|
Db      95 WTNWKF 100


RESULT 8
NIGM_HUMAN
ID   NIGM_HUMAN     STANDARD;      PRT;   105 AA.
AC   O95178;
DT   15-JUL-1999 (Rel. 38, Created)
DT   15-JUL-1999 (Rel. 38, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   NADH-ubiquinone oxidoreductase AGGG subunit, mitochondrial precursor
DE   (EC 1.6.5.3) (EC 1.6.99.3) (Complex I-AGGG) (CI-AGGG).
GN   NDUFB2.
OS   Homo sapiens (Human)
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20030858; PubMed=11042152;
RA   Loeffen J.L.C.M., Triepels R.H., van den Heuvel L., Schuelke M.,
RA   Buskens C.A.F., Smeets R.J.P., Trijbels J.M.F., Smeitink J.A.M.,
RT   "cDNA of eight nuclear encoded subunits of NADH:ubiquinone
RT   oxidoreductase: human complex I cDNA characterization completed.";
RL   Biochem. Biophys. Res. Commun. 273:415-422(2000).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Blood;
RX   MEDLINE=20489036; PubMed=11042152;
RA   Zhang Q.-F., Yi M., Wu X.-Y., Pan C.-X., Zhao M., Zhao C.-G., Fu G.,
RA   Shen Y., Fan H.-Y., Lu G., Zhong M., Xu X.-R., Han Z.-G., Zhang J.-W.,
RA   Tao J., Huang Q.-H., Zhou J., Hu G.-X., Gu J., Chen S.-J., Chen Z.;
RT   "Cloning and functional analysis of cDNAs with open reading frames for
RT   300 previously undetected genes expressed in CD34+ hematopoietic
RT   stem/progenitor cells.";
RL   Genome Res. 10:1546-1560(2000).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Lung;
RA   Strausberg R.;
RL   Submitted (DEC-2000) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: TRANSFER OF ELECTRONS FROM NADH TO THE RESPIRATORY
CC       CHAIN. THE IMMEDIATE ELECTRON ACCEPTOR FOR THE ENZYME IS BELIEVED
CC       TO BE UBIQUINONE.
CC   -!- CATALYTIC ACTIVITY: NADH + ubiquinone = NAD(+) + ubiquinol.
CC   -!- CATALYTIC ACTIVITY: NADH + H(+) + [ACP] = NAD(+) + [ACP]-S-acetyldihydrolipoamide.
CC   -!- SUBUNIT: COMPLEX I IS COMPOSED OF ABOUT 46 DIFFERENT SUBUNITS.
CC   -!- SUBCELLULAR LOCATION: Mitochondrial inner membrane, matrix side.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is not in any way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF050639; AAD05428.1; -.
DR   EMBL; AF067106; AAF20470.1; -.
DR   EMBL; BC001168; AAH01168.1; -.
DR   Genew; HGNC:7697; NDUFB2.
DR   MIM; 603838; -.
KW   Acetyltransferase; NAD; Ubiquinone; Mitochondrion; Transit peptide.
FT   TRANSIT      1     33       MITOCHONDRION (BY SIMILARITY).
FT   CHAIN       34    105       NADH-UBIQUINONE OXIDOREDUCTASE AGGG
FT                               SUBUNIT.
SQ   SEQUENCE   105 AA;  12058 MW;  5693366B6B7DC16F CRC64;

  Query Match             100.0%;  Score 31;  DB 1;  Length 105;
  Best Local Similarity   100.0%;  Pred. No. 1.2e+03;
  Matches    3;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 WXXWXF 6
           |::|:|
Db      68 WILWRF 73


RESULT 9
YQCC_HAEIN
ID   YQCC_HAEIN     STANDARD;      PRT;   106 AA.
AC   Q57152; Q05061;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Hypothetical protein HI1436.
GN   HI1436.
OS   Haemophilus influenzae.
OC   Bacteria; Proteobacteria, gamma subdivision, Pasteurellaceae;
OC   Haemophilus.
OX   NCBI_TaxID=727;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Rd / RWD / ATCC 51907;
RX   MEDLINE=95264990; PubMed=7542800;
RA   Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA   Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA   McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA   Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA   Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA   Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA   Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA   Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA   Venter J.C.;
RT   "Whole-genome random sequencing and assembly of Haemophilus influenzae
RT   Rd.";
RL   Science 269:496-512(1995).
CC   -!- SIMILARITY: TO THE N TERMINAL OF E.CARCTOVORA EXOENZYME REGULATION
CC       REGULON EXPI. THE C TERMINAL PART IS COLINEAR WITH YQCR.
CC   -!- SIMILARITY: STRONG, TO B.SUBTILI YQCC.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
```

```
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------------
DR   EMBL; U32822; AAC23085.1; -.
DR   TIGR; HI1436; -.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   106 AA;  12273 MW;  0955920EBD632280 CRC64;

  Query Match             100.0%;  Score 31;  DB 1;  Length 106;
  Best Local Similarity    50.0%;  Pred. No. 2.1e+02;
  Matches    3;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 WXXWXF 6
           |::|:|
Db      48 WLQWVF 53


RESULT 10
NIGM_BOVIN
ID   NIGM_BOVIN     STANDARD;      PRT;   108 AA.
AC   Q02374;
DT   01-JUL-1993 (Rel. 25, Created)
DT   01-JUL-1993 (Rel. 25, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   NADH ubiquinone oxidoreductase AGGG subunit, mitochondrial precursor
DE   (EC 1.6.5.3) (EC 1.6.99.3) (Complex I-AGGG) (CI-AGGG).
GN   NDUFB2.
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC   Bovidae; Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 32-65.
RC   TISSUE=Heart;
RX   MEDLINE=92380317; PubMed=1519044;
RA   Walker J.E., Arizmendi J.M., Dupuis A., Fearnley I.M., Finel M.,
RA   Medd S.M., Pilkington S.J., Runswick M.J., Skehel J.M.;
RT   "Sequences of 20 subunits of NADH:ubiquinone oxidoreductase from
RT   bovine heart mitochondria. Application of a novel strategy for
RT   sequencing proteins using the polymerase chain reaction.";
RL   J. Mol. Biol. 226:1051-1072(1992).
CC   -!- FUNCTION: TRANSFER OF ELECTRONS FROM NADH TO THE RESPIRATORY
CC       CHAIN. THE IMMEDIATE ELECTRON ACCEPTOR FOR THE ENZYME IS BELIEVED
CC       TO BE UBIQUINONE.
CC   -!- CATALYTIC ACTIVITY: NADH + ubiquinone = NAD(+) + ubiquinol.
CC   -!- CATALYTIC ACTIVITY: NADH + acceptor = NAD(+) + reduced acceptor.
CC   -!- SUBUNIT: COMPLEX I IS COMPOSED OF ABOUT 40 DIFFERENT SUBUNITS.
CC   -!- SUBCELLULAR LOCATION: Mitochondrial inner membrane; matrix side.
CC   ----------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------------
DR   EMBL; X63216; CAA44901.1; -.
DR   PIR; S28241; S28241.
KW   Oxidoreductase; NAD; Ubiquinone; Mitochondrion; Transit peptide.
FT   TRANSIT       1     36       MITOCHONDRION
FT   CHAIN        37    108       NADH-UBIQUINONE OXIDOREDUCTASE AGGG
FT                                SUBUNIT.
SQ   SEQUENCE   108 AA;  12282 MW;  0F0AD9B6A4912077 CRC64;

  Query Match             100.0%;  Score 31;  DB 1;  Length 108;
  Best Local Similarity    50.0%;  Pred. No. 2.1e+02;
  Matches    3;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 WXXWXF 6
           |::|:|
Db      71 WILWRF 76


RESULT 11
HV05_CARAU
ID   HV05_CARAU     STANDARD;      PRT;   116 AA.
AC   P19181;
DT   01-NOV-1990 (Rel. 16, Created)
DT   01-NOV-1990 (Rel. 16, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   Ig heavy chain V region 5A precursor.
OS   Carassius auratus (Goldfish).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii, Neopterygii, Teleostei; Ostariophysi, Cypriniformes;
OC   Cyprinidae; Carassius.
OX   NCBI_TaxID=7957;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=88144476; PubMed=3125551;
RA   Wilson M.R., Middleton D., Warr G.W.;
RT   "Immunoglobulin heavy chain variable region gene evolution: structure
RT   and family relationships of two genes and a pseudogene in a teleost
RT   fish.";
RL   Proc. Natl. Acad. Sci. U.S.A. 85:1566-1570(1988).
DR   PIR; B28966; B28966.
DR   HSSP; P01772; 2FB4.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR003596; Ig_V.
DR   Pfam; PF00047; ig; 1.
DR   SMART; SM00406; IGv; 1.
KW   Immunoglobulin V region; Signal.
FT   SIGNAL        1     19
FT   CHAIN        20    116       IG HEAVY CHAIN V REGION 5A.
FT   DOMAIN       20     49       FRAMEWORK-1.
FT   DOMAIN       50     54       COMPLEMENTARITY DETERMINING-1.
FT   DOMAIN       55     68       FRAMEWORK-2.
FT   DOMAIN       69     84       COMPLEMENTARITY-DETERMINING 2.
FT   DOMAIN       85    116       FRAMEWORK-3.
FT   DISULFID     41    114       BY SIMILARITY.
FT   NON_TER     116    116
SQ   SEQUENCE   116 AA;  12900 MW;  9F22A6F2BFB99B13 CRC64;

  Query Match             100.0%;  Score 31;  DB 1;  Length 116;
  Best Local Similarity    50.0%;  Pred. No. 2.2e+02;
  Matches    3;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 WXXWXF 6
           |::|:|
Db       4 WLSWVF 9


RESULT 12
MERT_STAAU
ID   MERT_STAAU     STANDARD;      PRT;   128 AA.
AC   P08656;
DT   01-JAN-1988 (Rel. 06, Created)
DT   01 JAN-1988 (Rel. 06, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   Mercuric transport protein (Mercury ion transport protein).
GN   MERT.
OS   Staphylococcus aureus.
OG   Plasmid pI258.
OC   Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX   NCBI_TaxID=1280;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=87260937; PubMed=3037534;
RA   Laddaga R.A., Chu L., Misra T.K., Silver S.;
RT   "Nucleotide sequence and expression of the mercurial resistance
RT   operon from Staphylococcus aureus plasmid pI258.";
RL   Proc. Natl. Acad. Sci. U.S.A. 84:5106-5110(1987).
CC   -!- FUNCTION: INVOLVED IN MERCURIC TRANSPORT. PASSES A HG(2+) ION
CC       FROM THE PERIPLASMIC MERP PROTEIN TO THE MERCURIC REDUCTASE
```

```
CC       (MERA).
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collabor-
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; L29436; AAA98244.1; -.
DR   PIR; D29594; D29594.
KW   Transport; Mercuric resistance; Transmembrane; Mercury; Plasmid.
FT   TRANSMEM     28     48       POTENTIAL.
FT   TRANSMEM     71     91       POTENTIAL.
FT   METAL        47     47       HG(2+) (POTENTIAL).
FT   METAL        48     48       HG(2+) (POTENTIAL).
FT   METAL       114    114       HG(2+) (POTENTIAL).
FT   METAL       115    115       HG(2+) (POTENTIAL).
SQ   SEQUENCE   128 AA,  14182 MW,  C3974E2D4A8AFE91 CRC64;

  Query Match             100.0%;  Score 31;  DB 1;  Length 128;
  Best Local Similarity   50.0%;  Pred. No. 2.4e+02;
  Matches    3;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 WXXWXF 6
           |::|:|
Db      28 WGVWAF 33


RESULT 13
KV3H_HUMAN
ID   KV3H_HUMAN     STANDARD;      PRT;   109 AA.
AC   P04207;
DT   20-MAR-1987 (Rel. 01, Created)
DT   01-NOV-1990 (Rel. 16, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   Ig kappa chain V-III region CLL precursor (Fragment) (Fragment).
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=86177570; PubMed=3083417;
RA   Klein R.D., Sirge J., Fang S., Heilman J.S., Cantor C., Chen P.P.,
RA   Goldfien R., Carson D.A.;
RT   "Cloning and sequence determination of a human kappa-III family
RT   light-chain gene.";
RL   Proc. Natl. Acad. Sci. U.S.A. 83:2199-2199(1986).
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collabor-
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; M12740; AAA88892.1; -.
DR   PIR; A01898; K3HUCL.
DR   HSSP; P80362; 1WTL.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR003596; Ig_V.
DR   Pfam; PF00047; ig; 1.
DR   SMART; SM00406; IGv; 1.
KW   Immunoglobulin V region; Signal.
FT   SIGNAL        1     20
FT   CHAIN        21    129       IG KAPPA CHAIN V-III REGION CLL.
FT   DOMAIN       21     43       FRAMEWORK-1.
FT   DOMAIN       44     54       COMPLEMENTARITY-DETERMINING-1.
FT   DOMAIN       55     69       FRAMEWORK-2.
FT   DOMAIN       70     76       COMPLEMENTARITY DETERMINING-2.
FT   DOMAIN       77    108       FRAMEWORK-3.
FT   DOMAIN      109    118       COMPLEMENTARITY-DETERMINING-3.
FT   DOMAIN      119    129       J-REGION.
FT   DISULFID     43    106       BY SIMILARITY.
FT   NON_TER     129    129
SQ   SEQUENCE   129 AA,  14[illegible] MW,  [illegible] CRC64;

  Query Match             [illegible]%;  Score 31;  DB 1;  Length 129;
  Best Local Similarity   50.0%;  Pred. No. 2.4e+02;
  Matches    3;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 WXXWXF 6
           |::|:|
Db     114 WPPWTF 119


RESULT 14
YHW2_YEAST
ID   YHW2_YEAST     STANDARD;      PRT;   129 AA.
AC   P38857;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   [illegible] (Rel. 40, Last annotation update)
DE   Hypothetical 14.6 kD protein in PET134-CYT1 intergenic region.
GN   YHR162W.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288C / AB972;
RX   MEDLINE=94185005; PubMed=8139559;
RA   Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Dover J.,
RA   Du Z., Favello A., Fulton L., Gattung S., Geisel C., Kirsten J.,
RA   Kucaba T., Hillier L., Jier M., Johnston L., Langston Y.,
RA   Latreille P., Louis E.J., Macri C., Mardis E., Menezes S., Mouser L.,
RA   Nhan M., Rifkin L., Riles L., St Peter H., Trevaskis E., Vaughan K.,
RA   Vignati D., Wilcox L., Wohldman P., Waterston R., Wilson R.,
RA   Vaudin M.;
RT   "Complete nucleotide sequence of Saccharomyces cerevisiae chromosome
RT   VIII.";
RL   Science 265:2077-2082(1994).
CC   -!- SIMILARITY: BELONGS TO THE UPF0041 (0-44) FAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collabor-
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; U00027; AAB68009.1; -.
DR   PIR; S48902; S48902.
DR   [illegible]; [illegible]; YHR162W.
DR   [illegible]; IPR[illegible]; UPF0041.
DR   Pfam; PF03650; UPF0041; 1.
KW   Hypothetical protein.
SQ   SEQUENCE   129 AA,  14[illegible] MW,  [illegible] CRC64;

  Query Match             100.0%;  Score 31;  DB 1;  Length 129;
  Best Local Similarity   50.0%;  Pred. No. 2.4e+02;
  Matches    3;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 WXXWXF 6
           |::|:|
Db      64 WTRWSF 69


RESULT 15
YA91_MYCPN
```

```
ID   YA91_MYCPN     STANDARD;      PRT;   138 AA.
AC   P75602;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hypothetical protein MPN091 (R02_orf138).
GN   MPN091 OR MP064.
OS   Mycoplasma pneumoniae.
OC   Bacteria; Firmicutes; Mollicutes, Mycoplasmataceae, Mycoplasma.
OX   NCBI_TaxID=2104;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN ATCC 29342 / M129,
RX   MEDLINE=97105885; PubMed=8948633;
RA   Himmelreich R., Hilbert H., Plagens H., Pirkl E., Li B.-C.,
RA   Herrmann R.;
RT   "Complete sequence analysis of the genome of the bacterium Mycoplasma
RT   pneumoniae.";
RL   Nucleic Acids Res. 24:4420-4449(1996).
CC   -!- SIMILARITY: ALMOST IDENTICAL TO M.PNEUMONIAAE MPN413 AND MPN463.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AE000008; AAB95711.1; -.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   138 AA;  15763 MW;  B52380D428F42709 CRC64;

  Query Match             100.0%;  Score 31;  DB 1;  Length 138;
  Best Local Similarity    50.0%;  Pred. No. 2.6e+02;
  Matches    3;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 WXXWXF 6
             |::|:|
Db      36 WSGWSF 41


Search completed: February 20, 2003, 10:48:02
Job time : 7.66667 secs
```

GenCore version 5.1.3

OM protein - protein search, using sw model

Run on February 20, 2003, 10:45:10 ; Search time 23.3333 Seconds
(without alignments)
52.994 Million cell updates/sec

Title: US-09-912-414-9
Perfect score: 31
Sequence: 1 WXXWXF 6

Scoring table: BLOSUM62EX
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_21:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution

SUMMARIES

Result No	Score	% Query Match	Length	DB	ID	Description
1	31	100.0	28	6	O62821	O62821 bubalus bub
2	31	100.0	31	8	Q94NH2	Q94nh2 listeria s
3	31	100.0	36	15	Q905D7	Q905d7 human immun
4	31	100.0	48	11	Q63365	Q63365 rattus norv
5	31	100.0	49	11	Q08615	Q08615 mus musculu
6	31	100.0	50	4	Q9NRC0	Q9nrc0 homo sapien
7	31	100.0	59	16	Q9PH20	Q9ph20 xylella fas
8	31	100.0	61	16	O85619	O85619 escherichia
9	31	100.0	64	4	Q8WVM9	Q8wvm9 homo sapien
10	31	100.0	70	2	Q9EXE4	Q9exe4 lactobacill
11	31	100.0	71	17	Q9HST8	Q9hst8 halobacteri
12	31	100.0	74	2	Q50168	Q50168 mycobacteri
13	31	100.0	76	16	O68793	O68793 yersinia pe
14	31	100.0	77	2	Q9S4W3	Q9s4w3 escherichia
15	31	100.0	79	2	Q9WTF7	Q9wtf7 escherichia
16	31	100.0	79	16	Q9JMR3	Q9jmr3 escherichia
17	31	100.0	80	16	Q93GP6	Q93gp6 salmonella
18	31	100.0	83	16	Q8UEM9	Q8uem9 agrobacteri
19	31	100.0	85	2	Q9P5A9	Q9p5a9 escherichia
20	31	100.0	86	2	Q9Z3F9	Q9z3f9 escherichia
21	31	100.0	86	10	Q8W1M2	Q8w1m2 solanum cha
22	31	100.0	87	16	Q9CGP4	Q9cgp4 [illegible]
23	31	100.0	89	10	Q8W1L9	Q8w1l9 solanum pin
24	31	100.0	89	10	Q8W1L8	Q8w1l8 solanum pin
25	31	100.0	92	10	Q8W1M0	Q8w1m0 solanum bul
26	31	100.0	92	12	Q90099	Q90099 turkey rhin
27	31	100.0	93	10	Q8W1M1	Q8w1m1 solanum bul
28	31	100.0	93	16	Q9CBW2	Q9cbw2 [illegible]
29	31	100.0	95	4	Q9Y6C4	Q9y6c4 homo sapien
30	31	100.0	96	4	Q96QD1	Q96qd1 homo sapien
31	31	100.0	97	16	Q9K5V6	Q9k5v6 bacillus ha
32	31	100.0	99	10	Q94AN6	Q94an6 arabidopsis
33	31	100.0	99	16	Q9K9M7	Q9k9m7 bacillus ha
34	31	100.0	101	5	Q95RS1	Q95rs1 drosophila
35	31	100.0	101	10	Q9XJ62	Q9xj62 triticum ae
36	31	100.0	103	16	Q9KTL5	Q9ktl5 vibrio chol
37	31	100.0	103	16	Q9XXI0	Q9xxi0 ralstonia s
38	31	100.0	105	11	Q9CPU2	Q9cpu2 mus musculu
39	31	100.0	105	16	Q9CNF4	Q9cnf4 pasteurella
40	31	100.0	106	16	Q9PB50	Q9pb50 xylella fas
41	31	100.0	106	16	Q9K1V7	Q9k1v7 chlamydia p
42	31	100.0	106	16	Q9ZF79	Q9zf79 chlamydia p
43	31	100.0	107	2	O06658	O06658 rhizobium l
44	31	100.0	108	16	Q8YYP9	Q8yyp9 anabaena sp
45	31	100.0	108	16	Q8Y923	Q8y923 listeria mo

ALIGNMENTS

RESULT 1
O62821
ID O62821 PRELIMINARY; PRT; 28 AA.
AC O62821;
DT 01-AUG-1998 (TrEMBLrel. 07, Created)
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Lysozyme (Fragment).
OS Bubalus bubalis (Domestic water buffalo).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC Bovidae; Bovinae; Bubalus.
OX NCBI_TaxID=89462;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BLOOD;
RA Das P., Tiwari G., Garg L.C.;
RT "Characterisation of lysozyme encoding gene promoter in Bubalus
RT bubalis.";
RL Submitted (MAR-1998) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AJ225012; CAA12342.1; -.
FT NON_TER 28 28
SQ SEQUENCE 28 AA; 3160 MW; B8CE6F6AD7DDB29D CRC64;

Query Match 100.0%; Score 31; DB 6; Length 28;
Best Local Similarity 50.0%; Pred. No. 2.9e+02;
Matches 3; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 WXXWXF 6
|::|:|
Db 10 WSVWTF 15

RESULT 2
Q94NH2
ID Q94NH2 PRELIMINARY; PRT; 31 AA.
AC Q94NH2;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)

```
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Cytochrome oxidase subunit II (Fragment).
OS   Littorina saxatilis.
OG   Mitochondrion.
OC   Eukaryota; Metazoa; Mollusca; Gastropoda; Orthogastropoda;
OC   Mesogastropoda; Littorinoidea; Littorinidae; Littorina.
OX   NCBI_TaxID=31220;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Moran P., Rolan Alvarez E.;
RL   Submitted (SEP-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF416614; AAL17810.1; -.
DR   EMBL; AF416615; AAL17812.1; -.
DR   InterPro; IPR002429; Cyt_c_oxy_2
DR   Pfam; PF02790; COX2_TM; 1.
KW   Mitochondrion.
FT   NON_TER      31     31
SQ   SEQUENCE   31 AA;  3654 MW;  C1DCE2ABBE9EB675 CRC64;

  Query Match             100.0%;  Score 31;  DB 8;  Length 31;
  Best Local Similarity   100.0%;  Pred. No. 4.9e+03;
  Matches    5;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 WXXWXF 6
             |:::|
Db       4 WGQWGF 9


RESULT 3
Q905D7
ID   Q905D7      PRELIMINARY;      PRT;    36 AA.
AC   Q905D7;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   Envelope glycoprotein (Fragment).
OS   Human immunodeficiency virus type 1
OC   Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX   NCBI_TaxID=11676;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=97CG316;
RA   Taniguchi Y., Takehsia J., Bikandou B., Mboudjeka I.,
RA   N'Doumbou N'Kiliz M. Y., M'Pandi M., M'Pele P., Harada Y., Hayami M.,
RA   Ichimura H., Parra H.J.;
RT   "Molecular subtypes of HIV Type 1 found in the epidemic sequenced in
RT   Republic of Congo Brazzaville.";
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF410464; AAM10285.1; -.
FT   NON_TER      36     36
SQ   SEQUENCE   36 AA;  4434 MW;  4BEAAB4C[illegible]B7AFA [illegible];

  Query Match             100.0%;  Score 31;  DB 15;  Length 36;
  Best Local Similarity    50.0%;  Pred. No. 3.7e+03;
  Matches    3;  Conservative    3;  Mismatches    [illegible];  Indels    [illegible];  Gaps    [illegible];

Qy       1 WXXWXF 6
             |:::|
Db      16 WGIWIF 21


RESULT 4
Q63365
ID   Q63365      PRELIMINARY;      PRT;    48 AA.
AC   Q63365;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Transforming oncogene NEU (Fragment).
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia, Eutheria, Rodentia, Sciurognathi, Muridae, Murinae, Rattus.
OX   NCBI_TaxID=10116;
```

```
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE 21561737; PubMed 11133292;
RA   Suen T.C., Hung M.C.;
RT   "Multiple cis- and trans-acting elements involved in regulation of the
RT   neu gene.";
RL   Mol. Cell. Biol. 10:6306-6315(1990).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE 91172172; PubMed 1672439;
RA   Yan L.H., Hung M.C.;
RT   "Identification and characterization of a novel enhancer for the rat
RT   neu promoter.";
RL   Mol. Cell. Biol. 11:1676-1682(1991).
DR   EMBL; M61004; AAA41686.1; -.
KW   Oncogene, Transforming protein.
FT   NON_TER      48     48
SQ   SEQUENCE   48 AA;  4966 MW;  7ZAUE4B1E13F1129 CRC64;

  Query Match             100.0%;  Score 31;  DB 11;  Length 48;
  Best Local Similarity    50.0%;  Pred. No. 4.9e+03;
  Matches    3;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 WXXWXF 6
             |:::|
Db       9 WCRWGF 14


RESULT 5
Q08615
ID   Q08615      PRELIMINARY;      PRT;    49 AA.
AC   Q08615;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   [illegible] (TrEMBLrel. 12, Last annotation update)
DE   Avian ERYTHROBLASTOSIS oncogene B 2 (EC 2.7.1.112) (NEU oncogene)
DE   (P185) (Epidermal growth factor receptor related protein) (Fragment).
GN   ERBB2 OR NEU.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE 92226496; PubMed 1349855;
RA   White M.K.-A., Hung M.-C.;
RT   "Cloning and characterization of the mouse neu promoter.";
RL   Oncogene [illegible]:683(1992).
CC   -!- CATALYTIC ACTIVITY: ATP + A PROTEIN TYROSINE = ADP + PROTEIN
CC       TYROSINE PHOSPHATE.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC   -!- SIMILARITY: EQUIVALENT TO HUMAN CERB B 2/HER 2.
CC   -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
DR   EMBL; X60636; CAA42966.1; -.
DR   MGI; MGI:99[illegible], ErbB2.
KW   Transmembrane; Glycoprotein, Duplication, Receptor; Transferase;
KW   Tyrosine protein kinase, ATP binding, Phosphorylation, Oncogene.
FT   NON_TER      49     49
SQ   SEQUENCE   49 AA;  5[illegible] MW;  E[illegible]DCB[illegible]B4E CRC64;

  Query Match             100.0%;  Score 31;  DB 11;  Length 49;
  Best Local Similarity    50.0%;  Pred. No. 4.9e+03;
  Matches    3;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 WXXWXF 6
             |:::|
Db       6 WCRWGF 11


RESULT 6
Q9NRC0
ID   Q9NRC0      PRELIMINARY;      PRT;    50 AA.
AC   Q9NRC0;
```

```
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   FAM4A1 (Fragment).
GN   FAM4A1.
OS   Homo sapiens (Human)
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20349292; PubMed=10889047;
RA   Vincent J.B., Herbrick J.A., Gurling H.M.D., Bolton P.F., Roberts W.,
RA   Scherer S.W.;
RT   "Identification of a novel gene on chromosome 7q31 that is interrupted
RT   by a translocation breakpoint in an autistic individual.";
RL   Am. J. Hum. Genet. 67:510-514(2001).
DR   EMBL; AF234886; AAF85949.1; -.
FT   NON_TER      50      50
SQ   SEQUENCE   50 AA;  5941 MW;  5A5BB51E24DDBCC3 CRC64;

  Query Match             100.0%;  Score 31;  DB 4;  Length 50;
  Best Local Similarity   50.0%;  Pred. No. 5e+02;
  Matches    3;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 WXXWXF 6
          |::|:|
Db     24 WTVWFF 29


RESULT 7
Q9PH20
ID   Q9PH20      PRELIMINARY;      PRT;    59 AA.
AC   Q9PH20;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Hypothetical protein XF0126.
GN   XF0126.
OS   Xylella fastidiosa.
OC   Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
OC   Xylella.
OX   NCBI_TaxID=2371;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=9A5C;
RX   MEDLINE=20365717; PubMed=10910347;
RA   Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,
RA   Alvarenga R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,
RA   Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,
RA   Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M.,
RA   Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa Neto C.M.,
RA   Coutinho L.L., Cristofani M., Dias Neto E., Docena C., El-Dorry H.,
RA   Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,
RA   Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furlan L.R.,
RA   Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
RA   Ho P.L., Hoheisel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,
RA   Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,
RA   Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
RA   Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,
RA   Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,
RA   Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
RA   Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
RA   Nhani A. Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,
RA   de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
RA   Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,
RA   Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
RA   de Rosa V.E. Jr., de Sa R.G., Santelli P.V., Sawasaki H.E.,
RA   da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,
RA   da Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,
RA   de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tsuhako M.H.,
RA   Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
RA   Zago M.A., Zatz M., Meidanis J., Setubal J.C.;
RT   "The genome sequence of the plant pathogen Xylella fastidiosa.";
RL   Nature 406:151-159(2000).
DR   EMBL; AE003866; AAF83939.1; -.
DR   InterPro; IPR001230; Prenyl_site.
DR   PROSITE; PS00294; PRENYLATION, UNKNOWN_1.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   59 AA;  7108 MW;  EE7D7CF0[illegible] CRC64;

  Query Match             100.0%;  Score 31;  DB 16;  Length 59;
  Best Local Similarity   50.0%;  Pred. No. 5.8e+02;
  Matches    3;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 WXXWXF 6
          |::|:|
Db     54 WRCWLF 59


RESULT 8
O85619
ID   O85619      PRELIMINARY;      PRT;    61 AA.
AC   O85619;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   L0012 (Hypothetical 5.5 kDa protein).
GN   Z5095 OR ECS4544.
OS   Escherichia coli, and
OS   Escherichia coli O157:H7.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
OX   NCBI_TaxID=562, 83334;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   SPECIES=E.coli; STRAIN=EDL933;
RX   MEDLINE=98179885; PubMed=9672066;
RA   Perna N.T., Mayhew G.F., Posfai G., Elliott S., Donnenberg M.S.,
RA   Kaper J.B., Blattner F.R.;
RT   "Molecular evolution of a pathogenicity island from enterohemorrhagic
RT   escherichia coli O157:H7.";
RL   Infect. Immun. 66:3810-3817(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   SPECIES=E.coli O157:H7; STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX   MEDLINE=21074935; PubMed=11206551;
RA   Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA   Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA   Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA   Grotbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
RA   Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA   Welch R.A., Blattner F.R.;
RT   "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL   Nature 409:529-533(2001).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   SPECIES=E.coli O157:H7; STRAIN=O157:H7 / RIMD [illegible];
RX   MEDLINE=21156231; PubMed=11258796;
RA   Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA   Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA   Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA   Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT   "Complete genome sequence of enterohemorrhagic Escherichia coli
RT   O157:H7 and genomic comparison with a laboratory strain K-12.";
RL   DNA Res. 8:11-22(2001).
DR   EMBL; AF071034; AAC31491.1; -.
DR   EMBL; AE005594; AAG58810.1; ALT_INIT.
DR   EMBL; AP002566; BAB37967.1; -.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   61 AA;  7286 MW;  5B7F4811F9D58CD3 CRC64;

  Query Match             100.0%;  Score 31;  DB 16;  Length 61;
  Best Local Similarity   50.0%;  Pred. No. 6e+02;
  Matches    3;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy       1 WXXWXF 6
           |:::|
Db       3 WDRWRF 8


RESULT 9
Q8WVM9
ID   Q8WVM9      PRELIMINARY;      PRT;    64 AA.
AC   Q8WVM9;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Hypothetical 7.0 kDa protein (Fragment).
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BRAIN;
RA   Strausberg E.;
RL   Submitted (DEC-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC017733; AAH17733.1; -.
KW   Hypothetical protein.
FT   NON_TER       1       1
SQ   SEQUENCE   64 AA;  7187 MW;  [illegible] CRC64;

  Query Match             100.0%;  Score 31;  DB 4;  Length 64;
  Best Local Similarity   [illegible]  Pred. No. 6.2e+02;
  Matches    3;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 WXXWXF 6
           |:::|
Db      16 WDFWLF 21


RESULT 10
Q9EXE4
ID   Q9EXE4      PRELIMINARY;      PRT;    70 AA.
AC   Q9EXE4;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Putative thymidylate synthase (Fragment).
GN   THY.
OS   Lactobacillus sakei
OC   Bacteria; Firmicutes, Bacillus/Clostridium group, Low G+C Bacillales;
OC   Lactobacillaceae; Lactobacillus.
OX   NCBI_TaxID=1599;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=LSAK8;
RA   Basso A., Campanile C., Ferrara L., Damiani G.;
RT   "Sequencing and analysis of DNA fragments obtained by random amplified
RT   polymorphic DNA (RAPD) technique from Lactobacillus sakei strains.";
RL   Submitted (FEB-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AJ414611; CAC29646.1; -.
DR   HSSP; P00469; 1TDY.
DR   InterPro; IPR000398; Thymidylat_synt.
DR   Pfam; PF00303; thymid_synt; 1.
DR   ProDom; PD001180; Thymidylat_synt; 1.
FT   NON_TER       1       1
FT   NON_TER      70      70
SQ   SEQUENCE   70 AA;  8488 MW;  [illegible] CRC64;

  Query Match             100.0%;  Score 31;  DB 4;  Length 70;
  Best Local Similarity   50.0%;  Pred. No. 6.8e+02;
  Matches    3;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 WXXWXF 6
           |::|:|
Db      60 WDEWAF 65


RESULT 11
Q9HST8
ID   Q9HST8      PRELIMINARY;      PRT;    71 AA.
AC   Q9HST8;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   Vng0079h.
GN   VNG0079H.
OS   Halobacterium sp. (strain NRC-1).
OC   Archaea; Euryarchaeota; Halobacteria; Halobacteriales;
OC   Halobacteriaceae; Halobacterium.
OX   NCBI_TaxID=64091;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE 20504482; PubMed 11016950;
RA   Ng W.V., Kennedy S.P., Mahairas G.G., Berquist B., Pan M.,
RA   Shukla H.D., Lasky S.R., Baliga N.S., Thorsson V., Sbrogna J.,
RA   Swartzell S., Weir D., Hall J., Dahl T.A., Welti R., Goo Y.A.,
RA   Leithauser B., Keller K., Cruz R., Danson M.J., Hough D.W.,
RA   Maddocks D.G., Jablonski P.E., Krebs M.P., Angevine C.M., Dale H.,
RA   Isenbarger T.A., Peck R.F., Pohlschroder M., Spudich J.L., Jung K.-H.,
RA   Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,
RA   Ebhardt H., Lowe T.M., Liang P., Riley M., Hood L., DasSarma S.;
RT   "Genome sequence of Halobacterium species NRC-1.";
RL   Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).
DR   EMBL; AE004976; AAG18714.1; -.
KW   Complete proteome.
SQ   SEQUENCE   71 AA;  [illegible] MW;  [illegible] CRC64;

  Query Match             100.0%;  Score 31;  DB 17;  Length 71;
  Best Local Similarity   50.0%;  Pred. No. 6.9e+02;
  Matches    3;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 WXXWXF 6
           |:::|:|
Db      49 WWLWGF 54


RESULT 12
Q50168
ID   Q50168      PRELIMINARY;      PRT;    74 AA.
AC   Q50168;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1996 (TrEMBLrel. 01, Last annotation update)
DE   U296u.
OS   Mycobacterium leprae.
OC   Bacteria; Firmicutes, Actinobacteria, Actinobacteridae;
OC   Actinomycetales, Corynebacterineae, Mycobacteriaceae, Mycobacterium.
OX   NCBI_TaxID 1769;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Smith D.R.;
RL   Submitted (APR-1995) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Robison K.;
RL   Submitted (SEP-1996) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U15187; AAA63114.1; -.
SQ   SEQUENCE   74 AA;  8949 MW;  [illegible] CRC64;

  Query Match             [illegible]  Score 31;  DB 2;  Length 74;
  Best Local Similarity   50.0%;  Pred. No. 7.1e+02;
  Matches    3;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 WXXWXF 6
           |::|:|
Db      59 WVKWLF 64


RESULT 13
```

```
O68793
ID   O68793      PRELIMINARY;      PRT;    76 AA.
AC   O68793;
DT   01-AUG-1998 (TrEMBLrel. 07, Created)
DT   01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Hypothetical protein Y1091.
GN   Y1091 OR YPMT1.60C.
OS   Yersinia pestis.
OG   Plasmid pMT1 (pMT-1).
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Yersinia.
OX   NCBI_TaxID=632;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KIM5;
RX   MEDLINE=98422474; PubMed=9748454;
RA   Hu P., Elliott J., Mcready P., Skowronski E., Garnes J.,
RA   Kobayashi A., Brubaker R.R., Garcia E.;
RT   "Structural organization of virulence-associated plasmids of Yersinia
RT   pestis.";
RL   J. Bacteriol. 180:5192-5202(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KIM10+;
RX   MEDLINE=99043898; PubMed=9826348;
RA   Lindler L.E., Plano G.V., Burland V., Mayhew G.F., Blattner F.R.;
RT   "Complete DNA sequence and detailed analysis of the Yersinia pestis
RT   KIM5 plasmid encoding murine toxin and capsular antigen.";
RL   Infect. Immun. 66:5731-5742(1998).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CO-92 / BIOVAR ORIENTALIS; PLASMID=PMT1 (PMT 1);
RX   MEDLINE=21470413; PubMed=11586360;
RA   Parkhill J., Wren B.W., Thomson N.R., Titball R.W., Holden M.T.G.,
RA   Prentice M.B., Sebaihia M., James K.D., Churcher C., Mungall K.L.,
RA   Baker S., Basham D., Bentley S.D., Brooks K., Cerdeno-Tarraga A.M.,
RA   Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,
RA   Feltwell T., Hamlin N., Holroyd S., Jagels K., Karlyshev A.V.,
RA   Leather S., Moule S., Oyston P.C.F., Quail M., Rutherford K.,
RA   Simmonds M., Skelton J., Stevens K., Whitehead S., Barrell B.G.;
RT   "Genome sequence of Yersinia pestis, the causative agent of plague.";
RL   Nature 413:523-527(2001).
DR   EMBL; AF053947; AAC13243.1; -.
DR   EMBL; AF074611; AAC82749.1; -.
DR   EMBL; AL117211; CAB55242.1; -.
KW   Hypothetical protein; Plasmid; Complete proteome.
SQ   SEQUENCE   76 AA;  8807 MW;  7FF44937755036BB CRC64;

  Query Match             100.0%;  Score 31;  DB 16;  Length 76;
  Best Local Similarity    50.0%;  Pred. No. 7.3e+02;
  Matches    3;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 WXXWXF 6
           |::|:|
Db      50 WRTWNF 55


RESULT 14
Q9S4W3
ID   Q9S4W3      PRELIMINARY;      PRT;    77 AA.
AC   Q9S4W3;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   Hypothetical 9.1 kDa protein.
OS   Escherichia coli.
OG   Plasmid F.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
OX   NCBI_TaxID=562;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99296678; PubMed=10366527;
RA   Manwaring N.P., Skurray R.A., Firth N.;
RT   "Nucleotide sequence of the F plasmid leading region.";
RL   Plasmid 41:219-225(1999).
DR   EMBL; AF106329; AAD47188.1; -.
KW   Hypothetical protein; Plasmid.
SQ   SEQUENCE   77 AA;  9089 MW;  38245FF025787977A CRC64;

  Query Match             100.0%;  Score 31;  DB 2;  Length 77;
  Best Local Similarity    50.0%;  Pred. No. 7.4e+02;
  Matches    3;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 WXXWXF 6
           |::|:|
Db      53 WRTWNF 58


RESULT 15
Q9WTF7
ID   Q9WTF7      PRELIMINARY;      PRT;    79 AA.
AC   Q9WTF7;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   YdeA protein.
GN   YDEA.
OS   Escherichia coli.
OG   Plasmid R100.
OC   Bacteria, Proteobacteria, gamma subdivision, Enterobacteriaceae;
OC   Escherichia.
OX   NCBI_TaxID=562;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Sampei G., Mizobuchi K.;
PT   "Organization and diversification of plasmid genomes: complete
PT   nucleotide sequence of the R100 genome.";
RL   Submitted (MAY-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AP000342; BAA78827.1; -.
SQ   SEQUENCE   79 AA;  9338 MW;  71C6B48F5D7615BC 7FCF4;

  Query Match             100.0%;  Score 31;  DB 2;  Length 79;
  Best Local Similarity    50.0%;  Pred. No. 7.6e+02;
  Matches    3;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 WXXWXF 6
           |::|:|
Db      55 WRTWNF 60


Search completed: February 20, 2003, 10:49:20
Job time : 25.3333 secs
```

GenCore version 5.1.3

OM protein - protein search, using sw model

Run on: February 20, 2003, 11:21:07 , Search time 29.3333 Seconds
(without alignments)
27.256 Million cell updates/sec

Title: US-09-912-414-2
Perfect score: 6
Sequence: 1 WVPWHF 6

Scoring table: OLIGOPX
Gapop 60.0 , Gapext 60.0

Searched: 908470 seqs, 133250620 residues

Word size : 0

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : A_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	6	100.0	6	21	AAB01492	Peptide which bind
2	5	83.3	6	21	AAB01504	Peptide which bind
3	5	83.3	6	21	AAB01509	Peptide which bind
4	5	83.3	372	20	AAY08982	Human LPA receptor
5	5	83.3	373	20	AAY05493	Human PSP 24 prote
6	5	83.3	373	22	AAU00227	LPA receptor-relat
7	4	66.7	6	14	AAR37389	Peptide for treati
8	4	66.7	6	21	AAB01505	Peptide which bind
9	4	66.7	6	21	AAB01508	Peptide which bind
10	4	66.7	10	17	AAR86140	Anti-ELAM-1 bindin
11	4	66.7	10	17	AAR86145	Anti-ELAM-1 bindin
12	4	66.7	10	17	AAR86146	Anti-ELAM-1 bindin
13	4	66.7	10	19	AAW63963	ELAM-1 peptide mim
14	4	66.7	10	19	AAW63964	ELAM-1 peptide mim
15	4	66.7	10	19	AAW63958	ELAM-1 peptide mim
16	4	66.7	12	14	AAR36519	D32.39 antibody is
17	4	66.7	12	15	AAR56756	Random peptide #53
18	4	66.7	12	17	AAR91534	D32.39 monoclonal
19	4	66.7	12	18	AAW25286	Antibody D32.39 ep
20	4	66.7	12	20	AAW80380	Peptide eluted aft
21	4	66.7	14	23	ABB74383	Karyophilic peptid
22	4	66.7	27	22	AAC12264	Human polypeptide
23	4	66.7	31	22	ABB43615	Peptide #11121 enc
24	4	66.7	31	22	ABB26569	Protein #8568 enco
25	4	66.7	31	22	AAM64556	Human brain expres
26	4	66.7	31	22	AAM77359	Human bone marrow
27	4	66.7	31	22	AAM21282	Peptide #2716 enco
28	4	66.7	31	22	AAM37509	Peptide #11546 enc
29	4	66.7	31	23	ABG46373	Human peptide enco
30	4	66.7	35	22	AAM87677	Human immune/haema
31	4	66.7	37	22	AAO13641	Human polypeptide
32	4	66.7	38	22	ABB30966	Peptide #3617 enco
33	4	66.7	38	22	ABB36158	Peptide #3664 enco
34	4	66.7	38	22	ABB21544	Protein #3543 enco
35	4	66.7	38	22	AAM69327	Human bone marrow
36	4	66.7	38	22	AAO12011	Human polypeptide
37	4	66.7	38	22	AAM17157	Peptide #3591 enco
38	4	66.7	38	22	AAM29648	Peptide #3685 enco
39	4	66.7	38	22	AAM04851	Peptide #3533 enco
40	4	66.7	38	23	ABG38943	Human peptide enco
41	4	66.7	41	22	AAO13303	Human polypeptide
42	4	66.7	42	22	AAM73679	Human bone marrow
43	4	66.7	42	22	AAM33865	Peptide #3552 enco
44	4	66.7	48	22	ABB64316	Drosophila melanog
45	4	66.7	49	22	AAO11866	Human polypeptide

ALIGNMENTS

RESULT 1
AAB01492
ID AAB01492 standard, peptide, 6 AA.
XX
AC AAB01492;
XX
DT 08-NOV-2000 (first entry)
XX
DE Peptide which binds to transcription factor E2F 1 DNA binding domain.
XX
KW DNA binding; transcription factor, E2F, E2F-1, cell cycle; DP-1;
KW activation, transcription, apoptosis, proliferative disorder;
KW psoriasis; restenosis.
XX
OS Synthetic.
XX
PN WO200044771-A1.
XX
PD 03-AUG-2000.
XX
PF 26-JAN-2000, 2000WO GB00227.
XX
PR 26-JAN-1999; 99GB-0001710.
XX
PA (PROL-) PROLIFIX LTD.
XX
PI Mueller R, Kontermann RE, Montigiani S;
XX
DR WPI; 2000-532806/48.
XX
PT Peptides binding to the DNA binding domain of transcription factor E2F
PT and inhibiting cell cycle progression, useful for the treatment of
PT cancer

```
XX
PS   Claim 6; Page 2; 42pp; English.
XX
CC   Peptides which bind to the DNA binding domain of transcription
CC   factor E2F and inhibit cell cycle progression may be useful as
CC   research agents to investigate the interaction between E2F and DP 1,
CC   or the activation of transcription by E2F 1/DP 1 heterodimers. They
CC   may also be used for inducing apoptosis and/or cell cycle arrest in
CC   a cell, particularly for treatment of cancer or other proliferative
CC   disorders such as psoriasis and restenosis.
XX
SQ   Sequence   6 AA;

  Query Match             100.0%;  Score 6;  DB 21;  Length 6;
  Best Local Similarity   [illegible]
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 WVRWHF 6
           ||||||
Db       1 WVRWHF 6


RESULT 2
AAB01504
ID   AAB01504 standard; peptide; 6 AA
XX
AC   AAB01504;
XX
DT   08-NOV-2000  (first entry)
XX
DE   Peptide which binds to transcription factor E2F-1 DNA binding domain.
XX
KW   DNA binding, transcription factor, E2F, E2F 1, cell cycle, DP 1,
KW   activation, transcription, apoptosis, proliferative disorder,
KW   psoriasis, restenosis.
XX
OS   Synthetic.
XX
PN   WO200044771 A1.
XX
PD   03-AUG-2000.
XX
PF   26-JAN-2000; 2000WO-GB00227.
XX
PR   26 JAN 1999;   99GB-0001710.
XX
PA   (PROL-) PROLIFIX LTD.
XX
PI   Mueller F,  Kontermann RE,  Montigiani S;
XX
DR   WPI; 2000-502806/48.
XX
PT   Peptides binding to the DNA binding domain of transcription factor E2F
PT   and inhibiting cell cycle progression, useful for the treatment of
PT   cancer
XX
PS   Example; Page 26; 42pp; English.
XX
CC   Peptides which bind to the DNA binding domain of transcription
CC   factor E2F and inhibit cell cycle progression may be useful as
CC   research agents to investigate the interaction between E2F and DP 1,
CC   or the activation of transcription by E2F 1/DP 1 heterodimers. They
CC   may also be used for inducing apoptosis and/or cell cycle arrest in
CC   a cell, particularly for treatment of cancer or other proliferative
CC   disorders such as psoriasis and restenosis.
XX
SQ   Sequence   6 AA;

  Query Match             83.3%;  Score 5;  DB 21;  Length 6;
  Best Local Similarity   [illegible]
  Matches    5;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       2 VRWHF 6
           ||||
Db       2 VRWHF 6


RESULT 3
AAB01509
ID   AAB01509 standard; peptide; 6 AA.
XX
AC   AAB01509;
XX
DT   08-NOV-2000  (first entry)
XX
DE   Peptide which binds to transcription factor E2F 1 DNA binding domain.
XX
KW   DNA binding, transcription factor, E2F, E2F 1, cell cycle; DP-1;
KW   activation, transcription, apoptosis, proliferative disorder;
KW   psoriasis; restenosis.
XX
OS   Synthetic.
XX
PN   WO200044771-A1.
XX
PD   03-AUG-2000.
XX
PF   26-JAN-2000; 2000WO-GB00227.
XX
PR   26 JAN 1999;   99GB-0001710.
XX
PA   (PROL-) PROLIFIX LTD.
XX
PI   Mueller F,  Kontermann RE,  Montigiani S;
XX
DR   WPI; 2000-502806/48.
XX
PT   Peptides binding to the DNA binding domain of transcription factor E2F
PT   and inhibiting cell cycle progression, useful for the treatment of
PT   cancer
XX
PS   Example; Page 26; 42pp; English.
XX
CC   Peptides which bind to the DNA binding domain of transcription
CC   factor E2F and inhibit cell cycle progression may be useful as
CC   research agents to investigate the interaction between E2F and DP 1,
CC   or the activation of transcription by E2F 1/DP 1 heterodimers. They
CC   may also be used for inducing apoptosis and/or cell cycle arrest in
CC   a cell, particularly for treatment of cancer or other proliferative
CC   disorders such as psoriasis and restenosis.
XX
SQ   Sequence   6 AA;

  Query Match             83.3%;  Score 5;  DB 21;  Length 6;
  Best Local Similarity   100.0%;  Pred. No. 7.8e+05;
  Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 WVRWH 5
           |||||
Db       1 WVRWH 5


RESULT 4
AAY89982
ID   AAY89982 standard; Protein; 372 AA.
XX
AC   AAY89982;
XX
DT   20-AUG-1999  (first entry)
XX
DE   Human LPA receptor protein.
XX
KW   LPA receptor; human; lysophosphatidic acid; 1-acyl-glycerol-3-phosphate;
KW   [illegible]; platelet aggregation; [illegible]
KW   physical abnormality, behavioural abnormality, cell proliferation, skin;
KW   neurite retraction, platelet aggregation, smooth muscle contraction;
```

```
KW   tumour cell invasion, neurotransmitter release, chemotaxis, cancer;
KW   drug design; tumour; anti coagulant; vascular restenosis, arthritis;
KW   anti-inflammatory; wound healing; tissue regeneration; blood coagulation;
KW   nerve regeneration; osteoporosis; bone regeneration; cosmetic; scarring;
KW   ischaemic heart disease; myocardial infarction; angina pectoris; stroke;
KW   cerebral infarction; cerebral sequelae; subarachnoid haemorrhage;
KW   benign prostatic hyperplasia; diabetes mellitus, retinopathy, neuropathy,
KW   nephropathy; diabetic retinopathy.
XX
OS   Homo sapiens.
XX
PN   WO9929887-A1.
XX
PD   17-JUN-1999.
XX
PF   10-DEC-1998;   98WO-US26262.
XX
PR   10-DEC-1997;   97US-0987943.
XX
PA   (SYNA-) SYNAPTIC PHARM CORP.
XX
PI   Bard JA;
XX
DR   WPI; 1999-385620/32.
DR   N-PSDB; AAX78139.
XX
PT   Lysophosphatidic acid  receptor sequence
XX
PS   Claim 9; Fig 3A-B; 121pp; English.
XX
CC   This invention describes a novel nucleic acid (I) encoding a human
CC   lysophosphatidic acid (LPA; 1-acyl-glycerol-3-phosphate) receptor. Probes
CC   and antibodies which bind to the LAP receptor can be used for detecting
CC   expression of a mammalian LPA receptor. The probe can be used for
CC   detecting a predisposition to a disorder associated with LPA receptor
CC   activity. The transgenic animals incorporating the LPA receptor can be
CC   used for determining the physiological effects of varying levels of
CC   activity of mammalian LPA receptors, or for identifying an (ant)agonist
CC   capable of alleviating a physical or behavioural abnormality associated
CC   with activity of LPA receptor. LPA receptors produce a wide range of
CC   responses such as cell proliferation, stimulation of neurite retraction,
CC   platelet aggregation, smooth muscle contraction, tumour cell invasion,
CC   neurotransmitter release, chloride efflux and chemotaxis. The human LPA
CC   receptor receptor can serve as a tool for the drug design of novel
CC   therapeutic agents for various indications, including cancer, tumour
CC   reduction or prevention), platelet aggregation (as an anti-coagulant),
CC   vascular restenosis, arthritis (as an anti-inflammatory), wound healing,
CC   tissue regeneration (preferably skin and nerve regeneration), blood
CC   coagulation, osteoporosis (bone regeneration), and cosmetic uses
CC   (preferably the prevention of abnormal growths or scarring or for
CC   augmentation). This receptor can additionally serve as a tool for the
CC   drug design of novel therapeutic agents for indications including
CC   ischaemic heart diseases (e.g. myocardial infarction and angina
CC   pectoris), stroke due to cerebrovascular disturbances (e.g. cerebral
CC   infarction and cerebral sequelae after subarachnoid haemorrhage), benign
CC   prostatic hyperplasia, complications of diabetes mellitus (e.g.
CC   retinopathy, neuropathy, and nephropathy), and diabetic retinopathy.
XX
SQ   Sequence   372 AA;

  Query Match             83.3%;  Score 5;  DB 20;  Length 372;
  Best Local Similarity   100.0%;  Pred. No. 8;
  Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      2 VRWHF 6
          |||||
Db     98 VRWHF 102


RESULT 5
AAY05493
ID   AAY05493 standard; Protein; 373 AA.
XX
AC   AAY05493;
XX
DT   07-JUL-1999  (first entry)
XX
DE   Human PSP-24 protein sequence.
XX
KW   EDG-1, EDG-2, EDG-3, EDG-4, EDG-5, PSP-24, human, detection, therapy;
KW   inverse agonist, allosteric modulator, lysophosphatidic acid receptor;
KW   LPA signalling mediated disease; cellular apoptosis.
XX
OS   Homo sapiens.
XX
PN   WO9919513-A2.
XX
PD   22-APR-1999.
XX
PF   09-OCT-1998;   98WO-US21315.
XX
PR   10-OCT-1997;   97US-0061572.
XX
PA   (LXRB-) LXR BIOTECHNOLOGY INC.
XX
PI   Erikson J,  Goddard JG,  Kiefer M;
XX
DR   WPI; 1999-277658/23.
DR   N-PSDB; AAX36570.
XX
PT   Identification of (ant)agonists of LPA receptor EDG-2 for, e.g.
PT   treating LPA signalling mediated diseases such as cellular apoptosis
XX
PS   Claim 1; Page 58-60; 63pp; English.
XX
CC   This sequence represents human PSP-24.
CC   The invention relates to methods of detecting (ant)agonist, inverse
CC   agonist or allosteric modulators of the lysophosphatidic acid receptors
CC   EDG-1, EDG-2, EDG-3, EDG-4, EDG-5, and PSP-24. The methods are used to
CC   identify (ant)agonists and allosteric modulator of the lysophosphatidic
CC   acid (LPA) EDG2 receptor, e.g. to treat LPA signalling mediated disease
CC   such cellular apoptosis.
XX
SQ   Sequence   373 AA;

  Query Match             83.3%;  Score 5;  DB 20;  Length 373;
  Best Local Similarity   100.0%;  Pred. No. 8;
  Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      2 VRWHF 6
          |||||
Db     98 VRWHF 102


RESULT 6
AAU00227
ID   AAU00227 standard; Protein; 373 AA.
XX
AC   AAU00227;
XX
DT   11-MAY-2001  (first entry)
XX
DE   LPA receptor-related amino acid sequence #1.
XX
KW   LPA receptor, EDG-2, lysophosphatidic acid, phospholipid, tumour;
KW   cell signalling, MAP kinase; LPA modulator; neurodegenerative disease;
KW   Alzheimer's disease; Parkinson's disease; neuron damage; apoptosis;
KW   ischaemic heart disease; viral; HIV; inflammatory bowel disease;
KW   organ transplant; mouse.
XX
OS   Mus sp.
XX
PN   WO200112838-A2.
XX
PD   22-FEB-2001.
XX
```

PF 11-AUG-2000; 2000WO-GB00701.
XX
PR 18-AUG-1999; 99US-0376399.
XX
PA (ATAI-) ATAIRGIN TECHNOLOGIES INC.
XX
PI Eric[illegible] J, Bland CE, Pfeifer M, Pfeifer T,
XX
DR WPI; 2001-225555/23.
DR N-PSDB; AAS00262.
XX
PT Modulating activity of a lysophosphatidic acid or its receptor for
PT treating tumours, viral diseases, involves interfering in LPA or its
PI receptor [illegible] binding site, eg with a phosphate group and
PI -
XX
PS Disclosure; Page 85 86; 89pp; English.
XX
CC The sequence represents the LPA receptor related protein 1 (EDG 2) used during analysis of lysophosphatidic acid receptor (LPA), EDG 2. The sequence is given in the specification but no further information is given. LPA is a phospholipid found in a variety of plant and animal products. EDG 2 is involved in cell signalling through activation of a MAP kinase cascade. Peptides represent modulating the activity of a lysophosphatidic acid (LPA) or LPA receptor (EDG 2) involves introducing a composition comprising LPA modulators to the LPA or receptor. The method is useful for treating diseases characterised by slowed growth or repair of neuronal cells, neurodegenerative diseases, such as Alzheimer's disease, Parkinson's disease, and acute neuron damage, for modulating apoptotic pathways and treating ischaemic heart disease, tumours, viral diseases bowel disease, and rejection of organ transplants.
XX
SQ Sequence [illegible] AA;

Query Match 83.3%; Score 5; DB 22; Length 373;
Best Local Similarity 100.0%; Pred. No. 8;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 VRWHF 6
||||
Db 98 VRWHF 102

RESULT 7
AAR37389
ID AAR37389 standard; peptide; 6 AA.
XX
AC AAR37389;
XX
DT 07-JUL-1993 (first entry)
XX
DE Peptide for treating septic shock.
XX
KW Toxic shock; blood endotoxin removal, serum, diagnostic reagent,
KW cytokine release inhibitor, treatment, pertussis, bacterial meningitis,
KW HIV related infections; polymyxin B; Group II.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Region 1..3
FT /note= "repeat region"
FT Region 4..6
FT /note= "repeat region"
XX
PN ZA9200943-A.
XX
PD 25-NOV-1992.
XX
PF 10-FEB-1992; 92ZA-0000943.
XX
PR 11-FEB-1991; 91US-0653744.
XX

PA (PROP/) PROPHYLAXIS M.
XX
PI Porro M;
XX
DR WPI; 1993-004304/11.
XX
PT New peptide for treatment or prevention of toxic shock comprises
PT specified sequences of amino acid(s) and analogs
PT comprising sequences with conservative
XX
PS Example; Page 5; 39pp; English.
XX
CC The (Group II) peptide is an example of a generic peptide of formula
CC R-(Lys/Arg/His Phe/Tyr/Trp Leu/Ile/Val)n-R, where n = 1-100
CC and each R is H, an amino acid residue or a fatty acid residue.
CC The peptide is useful for treating or preventing septic shock,
CC mixing with polymyxin B to reduce the toxicity, removing
CC endotoxins from blood, sera or other fluids (in vitro or in
CC vivo), controlling release of cytokines induced by endotoxins;
CC as diagnostic reagents to detect and quantify toxins in blood
CC serum, preparing non-toxic antigenic complexes of lipid A or
CC lipopolysaccharide (LPS), and for treating pertussis, bacterial
CC meningitis and HIV related infections. The usual dose is 10-100
CC mg/kg/day, given parenterally. It binds to the same sites as
CC polymyxin B, i.e. it inhibits all the toxic effects of lipid A. It
CC has no antibiotic activity, does not lyse erythrocytes, has no
CC toxicity in mice when injected at 50mg/kg and is relatively unstable
CC against proteases.
XX
SQ Sequence 6 AA;

Query Match 66.7%; Score 4; DB 14; Length 6;
Best Local Similarity 100.0%; Pred. No. 7.8e+05;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 WVRW 4
||||
Db 2 WVRW 5

RESULT 8
AAB01505
ID AAB01505 standard; peptide; 6 AA.
XX
AC AAB01505;
XX
DT 08-NOV-2000 (first entry)
XX
DE Peptide which binds to transcription factor E2F 1 DNA binding domain.
XX
KW DNA binding; transcription factor; E2F, E2F-1, cell cycle; DP-1;
KW activation; transcription; apoptosis, proliferative disorder;
KW psoriasis, restenosis.
XX
OS Synthetic.
XX
PN WO200044771-A1.
XX
PD 03-AUG-2000.
XX
PF 26-JAN-2000; 2000WO-GB00227.
XX
PR 28-JAN-1999; 99GB-0001710.
XX
PA (PROL-) PROLIFIX LTD.
XX
PI Mueller R, Kontermann RE, Montigiani S;
XX
DR WPI; 2000-532806/48.
XX
PT Peptides binding to the DNA binding domain of transcription factor E2F
PT and inhibiting cell cycle progression, useful for the treatment of
PT cancer

```
XX
PS   Example; Page 26; 42pp; English.
XX
CC   Peptides which bind to the DNA binding domain of transcription
CC   factor E2F and inhibit cell cycle progression may be useful as
CC   research agents to investigate the interaction between E2F and DP-1,
CC   or the activation of transcription by E2F-1/DP-1 heterodimers. They
CC   may also be used for inducing apoptosis and/or cell cycle arrest in
CC   a cell, particularly for treatment of cancer or other proliferative
CC   disorders such as psoriasis and restenosis.
XX
SQ   Sequence   6 AA;

  Query Match             66.7%;  Score 4;  DB 21;  Length 6;
  Best Local Similarity   100.0%;  Pred. No. 7.8e+05;
  Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      3 RWHF 6
          ||||
Db      3 RWHF 6


RESULT 9
AAB01508
ID   AAB01508 standard; peptide; 6 AA.
XX
AC   AAB01508;
XX
DT   08-NOV-2000  (first entry)
XX
DE   Peptide which binds to transcription factor E2F-1 DNA binding domain
XX
KW   DNA binding, transcription factor, E2F, E2F 1, cell cycle, DP-1,
KW   activation; transcription; apoptosis; proliferative disorder;
KW   psoriasis; restenosis.
XX
OS   Synthetic.
XX
PN   WO200044771-A1.
XX
PD   03-AUG-2000.
XX
PF   26-JAN-2000; 2000WO-GB00227.
XX
PR   26-JAN-1999;  99GB-0001710.
XX
PA   (PROL-) PROLIFIX LTD.
XX
PI   Mueller R,  Kontermann RE,  Montigiani S;
XX
DR   WPI; 2000-532806/48.
XX
PT   Peptides binding to the DNA binding domain of transcription factor E2F
PT   and inhibiting cell cycle progression, useful for the treatment of
PT   cancer
XX
PS   Example; Page 26; 42pp; English
XX
CC   Peptides which bind to the DNA binding domain of transcription
CC   factor E2F and inhibit cell cycle progression may be useful as
CC   research agents to investigate the interaction between E2F and DP 1,
CC   or the activation of transcription by E2F 1/DP-1 heterodimers. They
CC   may also be used for inducing apoptosis and/or cell cycle arrest in
CC   a cell, particularly for treatment of cancer or other proliferative
CC   disorders such as psoriasis and restenosis.
XX
SQ   Sequence   6 AA;

  Query Match             66.7%;  Score 4;  DB 21;  Length 6;
  Best Local Similarity   100.0%;  Pred. No. 7.8e+05;
  Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 WVRW 4
          ||||
Db      1 WVRW 4


RESULT 10
AAR86140
ID   AAR86140 standard, peptide, 10 AA.
XX
AC   AAR86140;
XX
DT   26-JUN-1996  (first entry)
XX
DE   Anti-ELAM-1 binding peptide #117.
XX
KW   Peptide mimetic; endothelial leukocyte adhesion molecule, ELAM; selectin;
KW   receptor, leukocyte, vascular wall, endothelium, extravasation;
KW   inflammation; sialyl Lewis; cell surface glycoprotein; HL60 cell.
XX
OS   Synthetic.
XX
PN   WO9531210-A1.
XX
PD   23-NOV-1995.
XX
PF   11 MAY 1995;   95WO-US06315.
XX
PR   11-MAY-1994;   94US-0241054.
XX
PA   (AFFY-) AFFYMAX TECHNOLOGIES NV.
XX
PI   Barrett RW,  Cwirla SE,  Dower WJ,  Koller KJ,  Lee J;
PI   Martens CL, Ruhland-fritsch B;
XX
DR   WPI, 1996-010687/01.
XX
PT   New peptide(s) that bind to endothelial leukocyte adhesion molecule
PT   1 - useful for treating inflammation and other E-selectin mediated
PT   diseases
XX
PS   Disclosure; Page 17; 85pp; English.
XX
CC   Peptides AAR86024-R86236 are examples of peptides and their mimetics
CC   that bind to endothelial leukocyte adhesion molecule (ELAM) 1. This
CC   molecule is a member of the selectin family of receptors and is involved
CC   in binding of leukocytes to the vascular endothelial wall prior to
CC   extravasation of the leukocyte, e.g . to a site of inflammation.
CC   The peptides bind pref to E-selectin but may also bind L- or
CC   P-selectin, and can be used to treat conditions mediated by E-selectin,
CC   e.g. inflammatory conditions. The peptides have strong affinity for the
CC   selectin receptors and inhibit the binding of the sialyl Lewis (SLe x)
CC   part of cell surface glycoproteins to E-selectin. The peptide are
CC   small, generally less than 2 KD, have an IC50 of up to 100 micromole
CC   against binding of HL60 cells to ELAM 1, have one or more peptide
CC   linkages replaced by CH2NH(CO)NR, phosphonate, CH2SO2NR, CH2NR, CON(R6),
CC   or NHCONH linkages where R = H or a lower alkyl and R6 = a lower alkyl.
CC   The peptides may also have substituted N- and C-termini e.g.
CC   succinimide, N-benzyloxycarbonyl or N lower alkyl cpds.
XX
SQ   Sequence   10 AA;

  Query Match             66.7%;  Score 4;  DB 17;  Length 10;
  Best Local Similarity   100.0%;  Pred. No. 9.6;
  Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 WVRW 4
          ||||
Db      6 WVRW 9


RESULT 11
AAR86145
ID   AAR86145 standard; peptide; 10 AA.
XX
```

```
AC   AAR86145;
XX
DT   26-JUN-1996  (first entry)
XX
DE   Anti-ELAM-1 binding peptide #117.
XX
KW   Peptide mimetic, endothelial leukocyte adhesion molecule, ELAM, selectin,
KW   receptor; leukocyte, vascular wall, endothelium, extravasation,
KW   inflammation, sialyl Lewis, cell surface glycoprotein, HL60 cell.
XX
OS   Synthetic.
XX
FH   Key             Location/qualifiers
FT   Modified-site   10
FT                   /note= "contain amidated C-terminus"
XX
PN   WO9531210-A1.
XX
PD   23-NOV-1995.
XX
PF   11-MAY-1995;   95WO-US06315.
XX
PR   11-MAY-1994;   94US-0241054.
XX
PA   (APPY-) APPYMAX TECHNOLOGIES NV.
XX
PI   Barrett RW,  Cwirla SE,  Dower WJ,  Koller KJ,  Lee J;
PI   Martens CL, Ruhland-fritsch B;
XX
DR   WPI; 1996-010687/01.
XX
PT   New peptide(s) that bind to endothelial leukocyte adhesion molecule
PT   1 - useful for treating inflammation and other E-selectin mediated
PT   diseases
XX
PS   Disclosure; Page 17; 85pp; English.
XX
CC   Peptides AAR86004-R86036 are examples of peptides and their mimetics
CC   that bind to endothelial leukocyte adhesion molecule (ELAM)-1.  This
CC   molecule is a member of the selectin family of receptors and is involved
CC   in binding of leukocytes to the vascular endothelial wall prior to
CC   extravasation of the leukocyte, e.g. to a site of inflammation.
CC   The peptides bind pref. to E-selectin but may also bind L- or
CC   P-selectin, and can be used to treat conditions mediated by E-selectin,
CC   e.g. inflammatory conditions.  The peptides have strong affinity for the
CC   selectin receptors and inhibit the binding of the sialyl Lewis (SLe-x)
CC   part of cell surface glycoproteins to E-selectin.  The peptide are
CC   small, generally less than 2 kD, have an IC50 of up to 100 micromole
CC   against binding of HL60 cells to ELAM-1, have one or more peptide
CC   linkages replaced by CH2NH(CO)NR, phosphonate, CH2SCONR, CH2NR, CON(R6),
CC   or NHCONR linkages where R = H or a lower alkyl and R6 = a lower alkyl.
CC   The peptides may also have substituted N- and C-termini e.g.
CC   succinimide, N-benzyloxycarbonyl or N-lower alkyl cpds.
XX
SQ   Sequence   10 AA;

  Query Match             66.7%;  Score 4;  DB 17;  Length 10;
  Best Local Similarity   100.0%;  Pred. No. 9.6;
  Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 WVRW 4
           ||||
Db       6 WVRW 9


RESULT 12
AAR86146
ID   AAR86146 standard; peptide; 10 AA.
XX
AC   AAR86146;
XX
DT   26-JUN-1996  (first entry)
XX
DE   Anti-ELAM-1 binding peptide #123.
XX
KW   Peptide mimetic, endothelial leukocyte adhesion molecule, ELAM, selectin;
KW   receptor, leukocyte, vascular wall, endothelium, extravasation;
KW   inflammation; sialyl lewis; cell surface glycoprotein; HL60 cell.
XX
OS   Synthetic.
XX
PN   WO9531210-A1.
XX
PD   23-NOV-1995.
XX
PF   11-MAY-1995;   95WO-US06315.
XX
PR   11-MAY-1994;   94US-0241054.
XX
PA   (APPY-) APPYMAX TECHNOLOGIES NV.
XX
PI   Barrett RW,  Cwirla SE,  Dower WJ,  Koller KJ,  Lee J;
PI   Martens CL, Ruhland-fritsch B;
XX
DR   WPI; 1996-010687/01.
XX
PT   New peptide(s) that bind to endothelial leukocyte adhesion molecule
PT   1 - useful for treating inflammation and other E-selectin mediated
PT   diseases
XX
PS   Disclosure; Page 17; 85pp; English.
XX
CC   Peptides AAR86004-R86036 are examples of peptides and their mimetics
CC   that bind to endothelial leukocyte adhesion molecule (ELAM)-1.  This
CC   molecule is a member of the selectin family of receptors and is involved
CC   in binding of leukocytes to the vascular endothelial wall prior to
CC   extravasation of the leukocyte, e.g. to a site of inflammation.
CC   The peptides bind pref. to E selectin but may also bind L- or
CC   P-selectin, and can be used to treat conditions mediated by E-selectin,
CC   e.g. inflammatory conditions.  The peptides have strong affinity for the
CC   selectin receptors and inhibit the binding of the sialyl Lewis (SLe-x)
CC   part of cell surface glycoproteins to E-selectin.  The peptide are
CC   small, generally less than 2 kD, have an IC50 of up to 100 micromole
CC   against binding of HL60 cells to ELAM-1, have one or more peptide
CC   linkages replaced by CH2NH(CO)NR, phosphonate, CH2SCONR, CH2NR, CON(R6),
CC   or NHCONR linkages where R = H or a lower alkyl and R6 = a lower alkyl.
CC   The peptides may also have substituted N- and C-termini e.g.
CC   succinimide, N-benzyloxycarbonyl or N-lower alkyl cpds.
XX
SQ   Sequence   10 AA;

  Query Match             66.7%;  Score 4;  DB 17;  Length 10;
  Best Local Similarity   100.0%;  Pred. No. 9.6;
  Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 WVRW 4
           ||||
Db       6 WVRW 9


RESULT 13
AAW63963
ID   AAW63963 standard; peptide; 10 AA.
XX
AC   AAW63963;
XX
DT   02-OCT-1998  (first entry)
XX
DE   ELAM-1 peptide mimetic #118.
XX
KW   Endothelial leukocyte adhesion molecule 1, ELAM-1; inflammation;
KW   selectin; diagnosis; mimetic.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
```

```
FT   Modified-site   10
FT                   /note= "C-terminal Met is amidated"
XX
PN   US5728802-A.
XX
PD   17-MAR-1998.
XX
PF   12-MAY-1995;   92US-0439817.
XX
PR   12-MAY-1995;   95US-0439817.
PR   06-MAY-1992;   92US-0881395.
PR   05-MAY-1993;   93US-0057295.
PR   11-MAY-1994;   94US-0241054
XX
PA   (AFFY-) AFFYMAX TECHNOLOGIES NV.
XX
PI   Barrett RW,  Cwirla SE,  Dower WJ,  Koller KJ,  Lee J;
PI   Martens CL,  Ruhland B;
XX
DR   WPI; 1998-249882/22.
XX
PT   Peptide(s) or their mimetic(s) that bind to E-selectin - useful for,
PT   e.g. treating conditions mediated by E-selectin such as inflammatory
PT   condition(s)
XX
PS   Example 2; Column 93-94; 84pp; English.
XX
CC   AAW63846-W64054 are peptides and peptide mimetics that bind selectins
CC   including endothelial leukocyte adhesion molecule 1 (ELAM-1) and can be
CC   used for blocking adhesion of leukocytes to the selectins. The peptides
CC   have applications for the treatment of conditions mediated by
CC   E-selectin, e.g. inflammatory conditions. They can also be used for
CC   diagnostic purposes, e.g. for identifying the vascular site of E-selectin
CC   in vivo or can be coupled to anti-inflammatory or other drugs.
XX
SQ   Sequence   10 AA;

  Query Match             66.7%;  Score 4;  DB 19;  Length 10;
  Best Local Similarity   100.0%;  Pred. No. 9.6;
  Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 WVRW 4
          ||||
Db      6 WVRW 9


RESULT 14
AAW63964
ID   AAW63964 standard; peptide; 10 AA.
XX
AC   AAW63964;
XX
DT   02-OCT-1998  (first entry)
XX
DE   ELAM-1 peptide mimetic #119.
XX
KW   Endothelial leukocyte adhesion molecule 1; ELAM-1; inflammation;
KW   selectin; diagnosis; mimetic.
XX
OS   Synthetic.
XX
PN   US5728802-A.
XX
PD   17-MAR-1998.
XX
PF   12-MAY-1995;   92US-0439817.
XX
PR   12-MAY-1995;   95US-0439817.
PR   06-MAY-1992;   92US-0881395.
PR   05-MAY-1993;   93US-0057295.
PR   11-MAY-1994;   94US-0241054.
XX
PA   (AFFY-) AFFYMAX TECHNOLOGIES NV.
XX
PI   Barrett RW,  Cwirla SE,  Dower WJ,  Koller KJ,  Lee J;
PI   Martens CL,  Ruhland B;
XX
DR   WPI; 1998-249882/22.
XX
PT   Peptide(s) or their mimetic(s) that bind to E-selectin - useful for,
PT   e.g. treating conditions mediated by E-selectin such as inflammatory
PT   condition(s)
XX
PS   Example 2; Column 93-94; 84pp; English.
XX
CC   AAW63846-W64054 are peptides and peptide mimetics that bind selectins
CC   including endothelial leukocyte adhesion molecule 1 (ELAM-1) and can be
CC   used for blocking adhesion of leukocytes to the selectins. The peptides
CC   have applications for the treatment of conditions mediated by
CC   E-selectin, e.g. inflammatory conditions. They can also be used for
CC   diagnostic purposes, e.g. for identifying the vascular site of E-selectin
CC   in vivo or can be coupled to anti-inflammatory or other drugs.
XX
SQ   Sequence   10 AA;

  Query Match             66.7%;  Score 4;  DB 19;  Length 10;
  Best Local Similarity   100.0%;  Pred. No. 9.6;
  Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 WVRW 4
          ||||
Db      6 WVRW 9


RESULT 15
AAW63958
ID   AAW63958 standard; peptide; 10 AA.
XX
AC   AAW63958;
XX
DT   02-OCT-1998  (first entry)
XX
DE   ELAM-1 peptide mimetic #113.
XX
KW   Endothelial leukocyte adhesion molecule 1; ELAM-1; inflammation;
KW   selectin; diagnosis; mimetic.
XX
OS   Synthetic.
XX
PN   US5728802-A.
XX
PD   17-MAR-1998.
XX
PF   12-MAY-1995;   92US-0439817.
XX
PR   12-MAY-1995;   95US-0439817.
PR   06-MAY-1992;   92US-0881395.
PR   05-MAY-1993;   93US-0057295.
PR   11-MAY-1994;   94US-0241054.
XX
PA   (AFFY-) AFFYMAX TECHNOLOGIES NV.
XX
PI   Barrett RW,  Cwirla SE,  Dower WJ,  Koller KJ,  Lee J;
PI   Martens CL,  Ruhland B;
XX
DR   WPI; 1998-249882/22.
XX
PT   Peptide(s) or their mimetic(s) that bind to E-selectin - useful for,
PT   e.g. treating conditions mediated by E-selectin such as inflammatory
PT   condition(s)
XX
PS   Example 2; Column 91-92, 84pp; English.
XX
CC   AAW63846-W64054 are peptides and peptide mimetics that bind selectins
CC   including endothelial leukocyte adhesion molecule 1 (ELAM-1) and can be
CC   used for blocking adhesion of leukocytes to the selectins. The peptides
```

CC have applications for the treatment of conditions mediated by
CC E-selectin, e.g. inflammatory conditions. They can also be used for
CC diagnostic purposes, e.g. for identifying the vascular site of E-selectin
CC in vivo or can be coupled to anti-inflammatory or other drugs.
XX
SQ Sequence 10 AA;

Query Match 66.7%; Score 4; DB 19; Length 10;
Best Local Similarity 100.0%; Pred. No. 9.6;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      1 WVRW 4
          ||||
Db      6 WVRW 9
```

Search completed: February 20, 2003, 11:26:31
Job time : 30.3333 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 20, 2003, 11:26:38 ; Search time 17.6667 Seconds
(without alignments)
30.539 Million cell updates/sec

Title: US-09-912-414-2
Perfect score: 6
Sequence: 1 WVRWHF 6

Scoring table: OLIGO0X
Gapop 60.0 , Gapext 60.0

Searched: 468380 seqs, 89021735 residues

Word size : 0

Total number of hits satisfying chosen parameters: 468380

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Pending_Patents_AA_New:*
1: /cgn2_6/ptodata/1/paa/PCT_NEW_COMB.pep:*
2: /cgn2_6/ptodata/1/paa/US06_NEW_COMB.pep:*
3: /cgn2_6/ptodata/1/paa/US07_NEW_COMB.pep:*
4: /cgn2_6/ptodata/1/paa/US08_NEW_COMB.pep:*
5: /cgn2_6/ptodata/1/paa/US09_NEW_COMB.pep:*
6: /cgn2_6/ptodata/1/paa/US10_NEW_COMB.pep:*
7: /cgn2_6/ptodata/1/paa/US60_NEW_COMB.pep:*

Pred No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	5	83.3	372	6	US-10-225-567A-394	Sequence 394, App
2	4	66.7	20	1	PCT-US02-34021-373	Sequence 373, App
3	4	66.7	20	1	PCT-US02-34021-422	Sequence 422, App
4	4	66.7	38	6	US-10-203-139A-13934	Sequence 13934, A
5	4	66.7	52	1	PCT-US02-32727-5526	Sequence 5526, Ap
6	4	66.7	52	5	US-09-978-825-5526	Sequence 5526, Ap
7	4	66.7	52	6	US-10-057-498-5526	Sequence 5526, Ap
8	4	66.7	56	1	PCT-US02-32727-7189	Sequence 7189, Ap
9	4	66.7	56	5	US-09-978-825-7189	Sequence 7189, Ap
10	4	66.7	56	6	US-10-057-498-7189	Sequence 7189, Ap
11	4	66.7	58	1	PCT-US02-32727-11626	Sequence 11626, A
12	4	66.7	58	5	US-09-978-825-11626	Sequence 11626, A
13	4	66.7	58	6	US-10-057-498-11626	Sequence 11626, A
14	4	66.7	64	1	PCT-US02-32727-24772	Sequence 24772, A
15	4	66.7	64	5	US-09-978-825-24772	Sequence 24772, A
16	4	66.7	64	6	US-10-057-498-24772	Sequence 24772, A
17	4	66.7	65	1	PCT-US02-32727-22483	Sequence 22483, A
18	4	66.7	65	5	US-09-514-9497-6665	Sequence 6665, Ap
19	4	66.7	65	5	US-09-978-825-22483	Sequence 22483, A
20	4	66.7	65	6	US-10-057-498-22483	Sequence 22483, A
21	4	66.7	66	1	PCT-US02-32727-25628	Sequence 25628, A
22	4	66.7	66	5	US-09-978-825-25628	Sequence 25628, A
23	4	66.7	66	6	US-10-057-498-25628	Sequence 25628, A
24	4	66.7	67	5	US-09-950-084-6443	Sequence 6443, Ap
25	4	66.7	70	1	PCT-US02-32727-23166	Sequence 23166, A
26	4	66.7	70	5	US-09-978-825-23166	Sequence 23166, A
27	4	66.7	70	6	US-10-057-498-23166	Sequence 23166, A
28	4	66.7	74	1	PCT-US02-32727-24738	Sequence 24738, A
29	4	66.7	74	5	US-09-978-825-24738	Sequence 24738, A
30	4	66.7	74	6	US-10-057-498-24738	Sequence 24738, A
31	4	66.7	79	1	PCT-US02-32727-15180	Sequence 15180, A
32	4	66.7	79	5	US-09-978-825-15180	Sequence 15180, A
33	4	66.7	79	6	US-10-057-498-15180	Sequence 15180, A
34	4	66.7	82	5	US-09-134-0000-4041	Sequence 4041, Ap
35	4	66.7	82	5	US-09-134-0000C-4041	Sequence 4041, Ap
36	4	66.7	102	1	PCT-US02-32727-3760	Sequence 3760, Ap
37	4	66.7	102	5	US-09-978-825-3760	Sequence 3760, Ap
38	4	66.7	102	6	US-10-057-498-3760	Sequence 3760, Ap
39	4	66.7	107	1	PCT-US02-32727-15052	Sequence 15052, A
40	4	66.7	107	5	US-09-978-825-15052	Sequence 15052, A
41	4	66.7	107	6	US-10-057-498-15052	Sequence 15052, A
42	4	66.7	113	1	PCT-US02-32727-11060	Sequence 11060, A
43	4	66.7	113	5	US-09-978-825-11060	Sequence 11060, A
44	4	66.7	113	6	US-10-057-498-11060	Sequence 11060, A
45	4	66.7	122	6	US-10-264-237-1899	Sequence 1899, Ap

ALIGNMENTS

RESULT 1
US-10-225-567A-394
; Sequence 394, Application US/10225567A
; GENERAL INFORMATION:
; APPLICANT: LifeSpan Biosciences
; APPLICANT: Brown, Joseph P.
; APPLICANT: Burmer, Glenna C.
; APPLICANT: Roush, Christine L.
; TITLE OF INVENTION: ANTIGENIC PEPTIDES AND ANTIBODIES TO G PROTEIN COUPLED RECEPTORS
; FILE REFERENCE: 1920-4-4
; CURRENT APPLICATION NUMBER: US/10/225,567A
; CURRENT FILING DATE: 2001-12-19
; PRIOR APPLICATION NUMBER: 60/257,144
; PRIOR FILING DATE: 2000-12-19
; NUMBER OF SEQ ID NOS: 2392
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 394
; LENGTH: 372
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-225-567A-394

Query Match 83.3%; Score 5; DB 6; Length 372;
Best Local Similarity 100.0%; Pred. No. 53;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 VRWHF 6
|||||
Db 98 VRWHF 102

RESULT 2
PCT-US02-34021-373
; Sequence 373, Application PC/TUS0234021
; GENERAL INFORMATION:
; APPLICANT: DGI BioTechnologies, Inc.
; TITLE OF INVENTION: TARGET SPECIFIC SCREENING AND ITS USE FOR IDENTIFYING DRUGS THAT BIND
; FILE REFERENCE: 0508-4009PC
; CURRENT APPLICATION NUMBER: PCT/US02/34021
; CURRENT FILING DATE: 2002-10-24
; PRIOR APPLICATION NUMBER: 60/345,471
; PRIOR FILING DATE: 2001-10-24
; NUMBER OF SEQ ID NOS: 537
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 373
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Escherichia coli
; FEATURE:

```
;   NAME/KEY: MISC_FEATURE
;   OTHER INFORMATION: VEGFR1-2DF 3 B1
PCT-US02-34021-373

  Query Match             66.7%;  Score 4;  DB 1;  Length 20;
  Best Local Similarity   100.0%;  Pred. No. 88;
  Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 VRWH 5
           ||||
Db      11 VRWH 14


RESULT 3
PCT-US02-34021-422
; Sequence 422, Application PC/TUS0234021
; GENERAL INFORMATION:
;  APPLICANT: DGI BioTechnologies, Inc.
;  TITLE OF INVENTION: TARGET SPECIFIC SCREENING AND ITS USE FOR IDENTIFYING TARGET BIND
;  FILE REFERENCE: 2598-4009PC
;  CURRENT APPLICATION NUMBER: PCT/US02/34021
;  CURRENT FILING DATE:  2002-10-24
;  PRIOR APPLICATION NUMBER: 60/347,171
;  PRIOR FILING DATE: 2001-10-24
;  NUMBER OF SEQ ID NOS: 537
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 422
;   LENGTH: 20
;   TYPE: PRT
;   ORGANISM: Escherichia coli
;   FEATURE:
;   NAME/KEY: MISC_FEATURE
;   OTHER INFORMATION: VEGFR2-2DF-A112
PCT-US02-34021-422

  Query Match             66.7%;  Score 4;  DB 1;  Length 20;
  Best Local Similarity   100.0%;  Pred. No. 88;
  Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 VRWH 5
           ||||
Db      11 VRWH 14


RESULT 4
US-10-203-138A-13934
; Sequence 13934, Application US/10203138A
; GENERAL INFORMATION:
;  APPLICANT: Molecular Dynamics, Inc.
;  APPLICANT:  Penn, Sharron G
;  APPLICANT:  Rank, David R
;  APPLICANT:  Hanzel, David K.
;  APPLICANT:  Chen, Wensheng
;  TITLE OF INVENTION: HUMAN GENOME DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
;  TITLE OF INVENTION:  ANALYSIS OF GENE EXPRESSION IN BT 474
;  FILE REFERENCE: PB 0004 WO 8
;  CURRENT APPLICATION NUMBER: US/10/203,138A
;  CURRENT FILING DATE:  2002-08-02
;  PRIOR APPLICATION NUMBER: US 60/180,312
;  PRIOR FILING DATE: 04 February 2000 (04.02.00)
;  PRIOR APPLICATION NUMBER: US 60/207,444
;  PRIOR FILING DATE: 26 May 2000 (26.05.00)
;  PRIOR APPLICATION NUMBER: US 09/641,...
;  PRIOR FILING DATE: 03 August 2000 (03.08.00)
;  PRIOR APPLICATION NUMBER: GB 24263.6
;  PRIOR FILING DATE: 03 October 2000 (03.10.00)
;  PRIOR APPLICATION NUMBER: US 60/236,353
;  PRIOR FILING DATE: 27 September 2000 (27.09.00)
;  PRIOR APPLICATION NUMBER: US 60/234,687
;  PRIOR FILING DATE: 21 September 2000 (21.09.00)
;  PRIOR APPLICATION NUMBER: US 09/606,408
;  PRIOR FILING DATE: 30 June 2000 (30.06.00)
;  NUMBER OF SEQ ID NOS: 15438
;  SOFTWARE: Molecular Dynamics Sequence Listing Engine
; SEQ ID NO 13934
;   LENGTH: 38
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: MAP TO AC012143.1
;   FEATURE:
;   OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.7
US-10-203-138A-13934

  Query Match             66.7%;  Score 4;  DB 6;  Length 38;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+02;
  Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 VRWH 5
           ||||
Db       1 VRWH 4


RESULT 5
PCT-US02-32727-5526
; Sequence 5526, Application PC/TUS0232727
; GENERAL INFORMATION:
;  APPLICANT: Mitcham, Jennifer
;  APPLICANT:  Skeiky, Yasir
;  APPLICANT:  Persing, David
;  APPLICANT:  Bhatia, Ajay
;  APPLICANT:  Malcolmeuve, Jean Francois
;  APPLICANT:  Zhang, Yanni
;  APPLICANT:  Wang, Siqing
;  APPLICANT:  Jen, Shyian
;  APPLICANT:  Lodes, Michael
;  APPLICANT:  Benson, Darin
;  APPLICANT:  Jones, Robert
;  APPLICANT:  Carter, Darrick
;  APPLICANT:  Barth, Brenda
;  APPLICANT:  Douglass,John
;  TITLE OF INVENTION: Compositions and Methods for the Therapy and Diagnosis of Acnes V
;  FILE REFERENCE: 210121.514C1
;  CURRENT APPLICATION NUMBER: PCT/US02/32727
;  CURRENT FILING DATE:  2002-10-11
;  NUMBER OF SEQ ID NOS: 30992
; SEQ ID NO 5526
;   LENGTH: 52
;   TYPE: PRT
;   ORGANISM: Propioni acnes
PCT-US02-32727-5526

  Query Match             66.7%;  Score 4;  DB 1;  Length 52;
  Best Local Similarity   100.0%;  Pred. No. 1.3e+02;
  Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 VRWH 5
           ||||
Db      26 VRWH 29


RESULT 6
PCT-US02-32727-5526
; Sequence 5526, Application PC/TUS0232727
; GENERAL INFORMATION:
;  APPLICANT: Mitcham, Jennifer
;  APPLICANT:  Skeiky, Yasir
;  APPLICANT:  Persing, David
;  APPLICANT:  Bhatia, Ajay
;  APPLICANT:  Malcolmeuve, Jean Francois
;  APPLICANT:  Zhang, Yanni
;  APPLICANT:  Wang, Siqing
;  APPLICANT:  Jen, Shyian
;  APPLICANT:  Lodes, Michael
;  APPLICANT:  Benson, Darin
;  APPLICANT:  Jones, Robert
```

```
;  APPLICANT:  Carter, Darrick
;  APPLICANT:  Barth, Brenda
;  APPLICANT:  Douglass,John
;  TITLE OF INVENTION: Compositions and Methods for the Therapy and Diagnosis of Acnes V
;  FILE REFERENCE: 210121.514C1
;  CURRENT APPLICATION NUMBER: US/09/978,825
;  CURRENT FILING DATE:  2003-01-29
;  NUMBER OF SEQ ID NOS: 30992
; SEQ ID NO 5526
;   LENGTH: 52
;   TYPE: PRT
;   ORGANISM: Propioni acnes
US-09-978-825-5526

  Query Match             66.7%;  Score 4;  DB 5;  Length 52;
  Best Local Similarity   100.0%;  Pred. No. 1.9e+02;
  Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 VRWH 5
           ||||
Db      26 VRWH 29


RESULT 7
US-10-057-498-5526
; Sequence 5526, Application US/10057498
; GENERAL INFORMATION:
;  APPLICANT: Mitcham, Jennifer
;  APPLICANT:  Skeiky, Yasir
;  APPLICANT:  Persing, David
;  TITLE OF INVENTION: Compositions and Methods for the Therapy and Diagnosis of Acnes V
;  FILE REFERENCE: 210121.514
;  CURRENT APPLICATION NUMBER: US/10/057,498
;  CURRENT FILING DATE:  2001-04-20
;  NUMBER OF SEQ ID NOS: 29212
; SEQ ID NO 5526
;   LENGTH: 52
;   TYPE: PRT
;   ORGANISM: Propioni acnes
US-10-057-498-5526

  Query Match             66.7%;  Score 4;  DB 6;  Length 52;
  Best Local Similarity   100.0%;  Pred. No. 1.9e+02;
  Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 VRWH 5
           ||||
Db      26 VRWH 29


RESULT 8
PCT-US02-32727-7189
; Sequence 7189, Application PC/TUS0232727
; GENERAL INFORMATION:
;  APPLICANT: Mitcham, Jennifer
;  APPLICANT:  Skeiky, Yasir
;  APPLICANT:  Persing, David
;  APPLICANT:  Bhatia, Ajay
;  APPLICANT:  Maisonneuve, Jean Francois
;  APPLICANT:  Zhang, Yanni
;  APPLICANT:  Wang, Siqing
;  APPLICANT:  Jen, Shyian
;  APPLICANT:  Lodes, Michael
;  APPLICANT:  Benson, Darin
;  APPLICANT:  Jones, Robert
;  APPLICANT:  Carter, Darrick
;  APPLICANT:  Barth, Brenda
;  APPLICANT:  Douglass,John
;  TITLE OF INVENTION: Compositions and Methods for the Therapy and Diagnosis of Acnes V
;  FILE REFERENCE: 210121.514C1
;  CURRENT APPLICATION NUMBER: PCT/US02/32727
;  CURRENT FILING DATE:  2002-10-11
;  NUMBER OF SEQ ID NOS: 30992
; SEQ ID NO 7189
;   LENGTH: 56
;   TYPE: PRT
;   ORGANISM: Propioni acnes
PCT-US02-32727-7189

  Query Match             66.7%;  Score 4;  DB 1;  Length 56;
  Best Local Similarity   100.0%;  Pred. No. 2e+02;
  Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 VRWH 5
           ||||
Db      20 VRWH 23


RESULT 9
US-09-978-825-7189
; Sequence 7189, Application US/09978825
; GENERAL INFORMATION:
;  APPLICANT: Mitcham, Jennifer
;  APPLICANT:  Skeiky, Yasir
;  APPLICANT:  Persing, David
;  APPLICANT:  Bhatia, Ajay
;  APPLICANT:  Maisonneuve, Jean Francois
;  APPLICANT:  Zhang, Yanni
;  APPLICANT:  Wang, Siqing
;  APPLICANT:  Jen, Shyian
;  APPLICANT:  Lodes, Michael
;  APPLICANT:  Benson, Darin
;  APPLICANT:  Jones, Robert
;  APPLICANT:  Carter, Darrick
;  APPLICANT:  Barth, Brenda
;  APPLICANT:  Douglass,John
;  TITLE OF INVENTION: Compositions and Methods for the Therapy and Diagnosis of Acnes V
;  FILE REFERENCE: 210121.514C1
;  CURRENT APPLICATION NUMBER: US/09/978,825
;  CURRENT FILING DATE:  2003-01-29
;  NUMBER OF SEQ ID NOS: 30992
; SEQ ID NO 7189
;   LENGTH: 56
;   TYPE: PRT
;   ORGANISM: Propioni acnes
US-09-978-825-7189

  Query Match             66.7%;  Score 4;  DB 5;  Length 56;
  Best Local Similarity   100.0%;  Pred. No. 2e+02;
  Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 VRWH 5
           ||||
Db      20 VRWH 23


RESULT 10
US-10-057-498-7189
; Sequence 7189, Application US/10057498
; GENERAL INFORMATION:
;  APPLICANT: Mitcham, Jennifer
;  APPLICANT:  Skeiky, Yasir
;  APPLICANT:  Persing, David
;  TITLE OF INVENTION: Compositions and Methods for the Therapy and Diagnosis of Acnes V
;  FILE REFERENCE: 210121.514
;  CURRENT APPLICATION NUMBER: US/10/057,498
;  CURRENT FILING DATE:  2001-04-20
;  NUMBER OF SEQ ID NOS: 29212
; SEQ ID NO 7189
;   LENGTH: 56
;   TYPE: PRT
;   ORGANISM: Propioni acnes
US-10-057-498-7189

  Query Match             66.7%;  Score 4;  DB 6;  Length 56;
  Best Local Similarity   100.0%;  Pred. No. 2e+02;
```

```
  Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 VRWH 5
           ||||
Db      20 VRWH 23


RESULT 11
PCT-US02-32727-11626
; Sequence 11626, Application PC/TUS0232727
; GENERAL INFORMATION:
;  APPLICANT: Mitcham, Jennifer
;  APPLICANT:  Skeiky, Yasir
;  APPLICANT:  Persing, David
;  APPLICANT:  Bhatia, Ajay
;  APPLICANT:  Maisonneuve, Jean Francois
;  APPLICANT:  Zhang, Yanni
;  APPLICANT:  Wang, Siging
;  APPLICANT:  Jen, Chyian
;  APPLICANT:  Lodes, Michael
;  APPLICANT:  Benson, Darin
;  APPLICANT:  Jones, Robert
;  APPLICANT:  Carter, Darrick
;  APPLICANT:  Barth, Brenda
;  APPLICANT:  Douglass,John
;  TITLE OF INVENTION: Compositions and Methods for the Therapy and Diagnosis of Acnes V
;  FILE REFERENCE: 210121.514C1
;  CURRENT APPLICATION NUMBER: PCT/US02/32727
;  CURRENT FILING DATE:  2002-10-11
;  NUMBER OF SEQ ID NOS: 30992
; SEQ ID NO 11626
;   LENGTH: 58
;   TYPE: PRT
;   ORGANISM: Propioni acnes
PCT-US02-32727-11626

  Query Match             66.7%;  Score 4;  DB 1;  Length 58;
  Best Local Similarity   100.0%;  Pred. No. 2e+02;
  Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 VRWH 5
           ||||
Db      11 VRWH 14


RESULT 12
US-09-978-825-11626
; Sequence 11626, Application US/09978825
; GENERAL INFORMATION:
;  APPLICANT: Mitcham, Jennifer
;  APPLICANT:  Skeiky, Yasir
;  APPLICANT:  Persing, David
;  APPLICANT:  Bhatia, Ajay
;  APPLICANT:  Maisonneuve, Jean Francois
;  APPLICANT:  Zhang, Yanni
;  APPLICANT:  Wang, Siging
;  APPLICANT:  Jen, Chyian
;  APPLICANT:  Lodes, Michael
;  APPLICANT:  Benson, Darin
;  APPLICANT:  Jones, Robert
;  APPLICANT:  Carter, Darrick
;  APPLICANT:  Barth, Brenda
;  APPLICANT:  Douglass,John
;  TITLE OF INVENTION: Compositions and Methods for the Therapy and Diagnosis of Acnes V
;  FILE REFERENCE: 210121.514C1
;  CURRENT APPLICATION NUMBER: US/09/978,825
;  CURRENT FILING DATE:  2001-01-29
;  NUMBER OF SEQ ID NOS: 30992
; SEQ ID NO 11626
;   LENGTH: 58
;   TYPE: PRT
;   ORGANISM: Propioni acnes
US-09-978-825-11626

  Query Match             66.7%;  Score 4;  DB 5;  Length 58;
  Best Local Similarity   100.0%;  Pred. No. 2e+02;
  Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 VRWH 5
           ||||
Db      11 VRWH 14


RESULT 13
US-10-057-498-11626
; Sequence 11626, Application US/10057498
; GENERAL INFORMATION:
;  APPLICANT: Mitcham, Jennifer
;  APPLICANT:  Skeiky, Yasir
;  APPLICANT:  Persing, David
;  TITLE OF INVENTION: Compositions and Methods for the Therapy and Diagnosis of Acnes V
;  FILE REFERENCE: 210121.514
;  CURRENT APPLICATION NUMBER: US/10/057,498
;  CURRENT FILING DATE:  2001-04-20
;  NUMBER OF SEQ ID NOS: 23212
; SEQ ID NO 11626
;   LENGTH: 58
;   TYPE: PRT
;   ORGANISM: Propioni acnes
US-10-057-498-11626

  Query Match             66.7%;  Score 4;  DB 6;  Length 58;
  Best Local Similarity   100.0%;  Pred. No. 2e+02;
  Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 VRWH 5
           ||||
Db      11 VRWH 14


RESULT 14
PCT-US02-32727-24772
; Sequence 24772, Application PC/TUS0232727
; GENERAL INFORMATION:
;  APPLICANT: Mitcham, Jennifer
;  APPLICANT:  Skeiky, Yasir
;  APPLICANT:  Persing, David
;  APPLICANT:  Bhatia, Ajay
;  APPLICANT:  Maisonneuve, Jean Francois
;  APPLICANT:  Zhang, Yanni
;  APPLICANT:  Wang, Siging
;  APPLICANT:  Jen, Chyian
;  APPLICANT:  Lodes, Michael
;  APPLICANT:  Benson, Darin
;  APPLICANT:  Jones, Robert
;  APPLICANT:  Carter, Darrick
;  APPLICANT:  Barth, Brenda
;  APPLICANT:  Douglass,John
;  TITLE OF INVENTION: Compositions and Methods for the Therapy and Diagnosis of Acnes V
;  FILE REFERENCE: 210121.514C1
;  CURRENT APPLICATION NUMBER: PCT/US02/32727
;  CURRENT FILING DATE:  2002-10-11
;  NUMBER OF SEQ ID NOS: 30992
; SEQ ID NO 24772
;   LENGTH: 64
;   TYPE: PRT
;   ORGANISM: Propioni acnes
PCT-US02-32727-24772

  Query Match             66.7%;  Score 4;  DB 1;  Length 64;
  Best Local Similarity   100.0%;  Pred. No. 2.2e+02;
  Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 VRWH 5
           ||||
Db       2 VRWH 5
```

```
RESULT 15
US-09-978-825-24772
; Sequence 24772, Application US/09978825
; GENERAL INFORMATION:
;  APPLICANT: Mitcham, Jennifer
;  APPLICANT:  Skeiky, Yasir
;  APPLICANT:  Persing, David
;  APPLICANT:  Bhatia, Ajay
;  APPLICANT:  Maisonneuve, Jean Francois
;  APPLICANT:  Zhang, Yanni
;  APPLICANT:  Wang, Siqing
;  APPLICANT:  Jen, Shyian
;  APPLICANT:  Lodes, Michael
;  APPLICANT:  Benson, Darin
;  APPLICANT:  Jones, Robert
;  APPLICANT:  Carter, Darrick
;  APPLICANT:  Barth, Brenda
;  APPLICANT:  Douglass,John
;  TITLE OF INVENTION: Compositions and Methods for the Therapy and Diagnosis of Acnes V
;  FILE REFERENCE: 210121.514C1
;  CURRENT APPLICATION NUMBER: US/09/978,825
;  CURRENT FILING DATE:  2003-01-29
;  NUMBER OF SEQ ID NOS: 30992
; SEQ ID NO 24772
;   LENGTH: 64
;   TYPE: PRT
;   ORGANISM: Propioni acnes
US-09-978-825-24772

  Query Match             66.7%;  Score 4;  DB 5;  Length 64;
  Best Local Similarity   100.0%;  Pred. No. 2.2e+02;
  Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 VRWH 5
           ||||
Db       2 VRWH 5


Search completed: February 20, 2003, 11:42:48
Job time : 17.6667 secs
```

GenCore version 5.1.3

OM protein - protein search, using sw model

Run on: February 20, 2003, 11:24:02 ; Search time 12 Seconds
(without alignments)
48.067 Million cell updates/sec

Title: US-09-912-414-2
Perfect score: 6
Sequence: 1 WVPWHF 6

Scoring table: OLIGODX
Gapop 60.0 , Gapext 60.0

Searched: 283224 seqs, 90134422 residues

Word size : 0

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : PIR_73:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	5	83.3	301	2	G83352	transcription regu
2	5	83.3	370	2	A96741	hypothetical prote
3	5	83.3	799	2	T48889	serine/threonine p
4	4	66.7	52	2	A95009	hypothetical prote
5	4	66.7	52	2	E97880	hypothetical prote
6	4	66.7	62	2	JC6155	viral capsid prote
7	4	66.7	95	2	S10796	conserved hypothet
8	4	66.7	98	2	AG3064	hypothetical prote
9	4	66.7	102	2	A99800	hypothetical prote
10	4	66.7	103	2	T27486	hypothetical prote
11	4	66.7	106	2	T44499	hypothetical prote
12	4	66.7	119	1	MNXP1R	sigma 1bNS protein
13	4	66.7	119	2	P34829	sigma 1bNS protein
14	4	66.7	122	2	T04366	probable peroxidas
15	4	66.7	133	2	B97409	hypothetical prote
16	4	66.7	133	2	AC2618	conserved hypothet
17	4	66.7	135	2	DR1704	hypothetical prote
18	4	66.7	149	1	GGLMF	globin - river lam
19	4	66.7	149	1	GGLMS	globin V [validate
20	4	66.7	149	2	A26042	globin I - sea lam
21	4	66.7	149	2	B26042	globin III - sea l
22	4	66.7	149	2	S13458	hemoglobin - south
23	4	66.7	149	2	S13459	hemoglobin - south
24	4	66.7	149	2	S13460	hemoglobin - south
25	4	66.7	162	2	F82493	hypothetical prote
26	4	66.7	169	2	F75253	conserved hypothet
27	4	66.7	170	2	S64488	regulatory protein
28	4	66.7	171	2	AD2457	hypothetical prote
29	4	66.7	188	2	A71286	conserved hypothet
30	4	66.7	197	2	A92556	c-type cytochrome
31	4	66.7	206	2	T40138	hypothetical prote
32	4	66.7	208	2	H69359	unnamed hypothet
33	4	66.7	209	2	C97337	HD superfamily hyd
34	4	66.7	210	2	D84016	hypothetical prote
35	4	66.7	215	2	T36800	probable transcrip
36	4	66.7	221	2	D97161	probable enzyme wi
37	4	66.7	226	2	C86255	protein F12F1.11 [
38	4	66.7	232	2	H75422	probable polysacch
39	4	66.7	233	1	H71040	probable phosphoes
40	4	66.7	241	2	F72385	conserved hypothet
41	4	66.7	244	2	T15996	hypothetical prote
42	4	66.7	249	2	T12463	hypothetical prote
43	4	66.7	253	2	T01453	hypothetical prote
44	4	66.7	257	2	C69862	conserved hypothet
45	4	66.7	257	2	C83982	hypothetical prote

ALIGNMENTS

RESULT 1
G83352
transcription regulator MtlR PA2337 [imported] - Pseudomonas aeruginosa (strain PAO1)
C;Species: Pseudomonas aeruginosa
C;Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C;Accession: G83352
R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrener, P.; Hickey, M.J.; Brinkman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim, .; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A;Title: Complete genome sequence of Pseudomonas aeruginosa PAO1, an opportunistic pathoc
A;Reference number: A82950; MUID:20437337; PMID:10984043
A;Accession: G83352
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-301 <STO>
A;Cross-references: GB AE004481; GP AE004481; NID g9947233; PIDN AAF91127.1; GSPDB:GN00101
A;Experimental source: strain PAO1
C;Genetics:
A;Gene: mtlR; PA2337

Query Match 83.3%; Score 5; DB 2; Length 301;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 VRWHF 6
|||||
Db 35 VRWHF 39

RESULT 2
A96741
hypothetical protein F14O23.22 [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C;Accession: A96741
R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso, .Chin, C.W.; Chung, M.K.; Conn, T.; Conway, A.R.; Conway, A.R.; Dewar, K.; ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C. C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali, Fizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, L ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A;Reference number: A86141; MUID:21016713; PMID:11130712
A;Accession: A96741
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-370 <STO>
A;Cross-references: GB AE005173; NID g7239911; PIDN AAF43137.1; GSPDB:GN00141

C;Genetics:
A;Gene: F14G23.22
A;Map position: 1

  Query Match [illegible] Score 6; DB 2; Length [illegible];
  Best Local Similarity 100.0%; Pred. No. [illegible];
  Matches [illegible]; Conservative [illegible]; Mismatches [illegible]; Indels [illegible]; Gaps [illegible];

Qy 1 WVPWH 5
      |||||
Db 151 WVPWH 155

RESULT 3
T48889
serine/threonine protein kinase (EC 2.7.1.-) afsR [validated] - Streptomyces coelicolor
C;Species: Streptomyces coelicolor
C;Date: 02-Jun-2000 #sequence_revision 02-Jun-2000 #text_change 02-Jun-2000
C;Accession: T48889
R;Matsumoto, A.; Hong, S.; Ishizuka, H.; Horinouchi, S.; Beppu, T.
Gene 146, 47-56, 1994
A;Title: Phosphorylation of the AfsR protein involved in secondary metabolism in Streptom
A;Reference number: Z24852; MUID:94341568; PMID:8063104
A;Accession: T48889
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-799 <MAT>
A;Cross-references: EMBL:D15062; PIDN:BAA03661.1
A;Experimental source: strain M130
C;Genetics:
A;Gene: afsK
C;Function:
A;Description: EC 2.7.1.-; [validated, MUID:94341568]; phosphorylates AfsR, a global regu
C;Keywords: phosphotransferase

  Query Match 83.3%; Score 5; DB 2; Length 799;
  Best Local Similarity 100.0%; Pred. No. 26;
  Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 VPWWH 6
      |||||
Db 628 VPWWH 632

RESULT 4
A95009
hypothetical protein SP0069 [imported] - Streptococcus pneumoniae (strain TIGR4)
C;Species: Streptococcus pneumoniae
C;Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 03-Aug-2001
C;Accession: A95009
R;Tettelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heid
on, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapple,
nson, T.; Hickey, E.K.; Holt, I.E.
Science 293, 498-506, 2001
A;Authors: Tettelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson,
A;Title: Complete Genome Sequence of a Virulent Isolate of Streptococcus pneumoniae.
A;Reference number: A95000; MUID:21567204; PMID:11468916
A;Accession: A95009
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-52 <KUR>
A;Cross-references: GB:AE007372; PIDN:AAK74777.1; PID:g15901744; GSPDB:GN00174; GB:AF3
A;Experimental source: strain TIGR4
C;Genetics:
A;Gene: SP0069

  Query Match 66.7%; Score 4; DB 2; Length 52;
  Best Local Similarity 100.0%; Pred. No. 74;
  Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 RWHF 6
      ||||
Db 19 RWHF 22

RESULT 5
E97880
hypothetical protein spr0069 [imported] - Streptococcus pneumoniae (strain R6)
C;Species: Streptococcus pneumoniae
C;Date: 22-Jul-2001 #sequence_revision 22-Jul-2001 #text_change 22-Oct-2001
C;Accession: E97880
R;Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszczak, L.; Burgett, S.; DeHoff, B.S.; Es
e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, C.; Matsushima, P.; McAhren, S.; Mc
y, P.; Sun, P.M.; Winkler, M.E.
J. Bacteriol. 183, 5709-5717, 2001
A;Authors: Yang, Y.; Young Bullis, M.; Chu, G.; Dick, C.; Palys, P.W.; Jackman, J.R.;
A;Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.
A;Reference number: A97872; MUID:21420245; PMID:11544234
A;Accession: E97880
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-52 <KUR>
A;Cross-references: GB:AF003917; PIDN:AAK98873.1; PID:g15467603; GSPDB:GN01174
C;Genetics:
A;Gene: spr0069

  Query Match 66.7%; Score 4; DB 2; Length 52;
  Best Local Similarity 100.0%; Pred. No. 74;
  Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 RWHF 6
      ||||
Db 19 RWHF 22

RESULT 6
JC6155
viral capsid protein 62 - human herpesvirus 8
C;Species: human herpesvirus 8, Kaposi sarcoma-associated virus
C;Date: 02-Sep-1997 #sequence_revision 02-Sep-1997 #text_change 19-May-2000
C;Accession: JC6155
R;Zhong, W.; Wang, H.; Herndier, B.; Ganem, D.
Proc. Natl. Acad. Sci. U.S.A. 93, 6641-6646, 1996
A;Title: Restricted expression of Kaposi sarcoma-associated herpesvirus (human herpesvir
A;Reference number: JC6153; MUID:96370595; PMID:8692871
A;Accession: JC6155
A;Molecule type: mRNA
A;Residues: 1-62 <ZHO>
A;Cross-references: GB:U75698
A;Note: the authors translated the codon TTT for residue 16 as Glu
C;Comment: This protein is a marker for the subgroup 1 : involved in productive infectio
C;Keywords: capsid protein

  Query Match 66.7%; Score 4; DB 2; Length 62;
  Best Local Similarity 100.0%; Pred. No. 85;
  Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 VRWH 5
      ||||
Db 31 VRWH 34

RESULT 7
S10796
conserved hypothetical protein [imported] - Bacillus stearothermophilus (fragment)
C;Species: Bacillus stearothermophilus
C;Date: 21-Nov-1990 #sequence_revision 10-Nov-1995 #text_change 01-Feb-2000
C;Accession: S10796; T46887
R;Hawkins, C.F.; Borges, A.; Perham, R.N.
Eur. J. Biochem. 191, 337-346, 1990
A;Title: Cloning and sequence analysis of the genes encoding the alpha and beta subunits
A;Reference number: S10796; MUID:90345939; PMID:2200674
A;Accession: S10796
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-95 <HAW>

```
A;Cross-references: EMBL:X83560, NID g40038, PIDN CAA37626.1, PID g1334248
R;Borges, A.; Hawkins, C.F.; Packman, L.C.; Perham, R.N.
Eur. J. Biochem. 194, 95-102, 1990
A;Title: Cloning and sequence analysis of the genes encoding the dihydrolipoamide acetyl
Bacillus stearothermophilus.
A;Reference number: S13838; MUID:91071217; PMID:2253629
A;Accession: T46887
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-95 <BOR>
A;Cross-references: EMBL:X83560, NID g40038, PIDN:CAA37626.1; PID:g1334248
A;Experimental source: strain NCA1503
C;Superfamily: Methanobacterium thermoautotrophicum conserved hypothetical protein MTH10

  Query Match             66.7%;  Score 4;  DB 2;  Length 95;
  Best Local Similarity   100.0%;  Pred. No. 1.2e+02;
  Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 VRWH 5
           ||||
Db       5 VRWH 8


RESULT 8
AG3064
hypothetical protein Atu4133 [imported] - Agrobacterium tumefaciens (strain C58, Dupont)
C;Species: Agrobacterium tumefaciens
C;Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 11-Jan-2002
C;Accession: AG3064
R;Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Ward, G.E.; Chen, Y.; Woo, L
erage, G.; Giller, W.; Grant, C.; Guenthner, D.; Kutyavin, T.; Levy, R.; Li, M.; McClell
; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A;Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
ster, E.W.
A;Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58
A;Reference number: AB2577; PMID:11743193
A;Accession: AG3064
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-98 <KUR>
A;Cross-references: GB:AE008690; PIDN:AAL44933.1; PID:g17742593; GSPDB:GN00187
A;Experimental source: strain C58 (Dupont)
C;Genetics:
A;Gene: Atu4133
A;Map position: linear chromosome

  Query Match             66.7%;  Score 4;  DB 2;  Length 98;
  Best Local Similarity   100.0%;  Pred. No. 1.2e+02;
  Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 VRWH 5
           ||||
Db       3 VRWH 6


RESULT 9
A99800
hypothetical protein ECs1369 [imported] - Escherichia coli (strain O157:H7, substrain RI
C;Species: Escherichia coli
C;Date: 19-Jul-2001 #sequence_revision 19-Jul-2001 #text_change 19-Jul-2001
C;Accession: A99800
R;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A;Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genc
A;Reference number: A99620; MUID:21156231; PMID:11258796
A;Accession: A99800
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-102 <HAY>
A;Cross-references: GB:BA000007; PIDN:BAB34792.1; PID:g13360829; GSPDB:GN00154
A;Experimental source: strain O157:H7, substrain RIMD 0509952
C;Genetics:
A;Gene: ECs1369

  Query Match             66.7%;  Score 4;  DB 2;  Length 102;
  Best Local Similarity   100.0%;  Pred. No. 1.3e+02;
  Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 RWHF 6
           ||||
Db      39 RWHF 42


RESULT 10
T27486
hypothetical protein Y9C2UA.2 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 04-Mar-2000
C;Accession: T27486
R;Smye, R.
submitted to the EMBL Data Library, December 1998
A;Reference number: Z20374
A;Accession: T27486
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-103 <WIL>
A;Cross-references: EMBL:AL009841; PIDN:CAA[illegible]; [illegible]
A;Experimental source: clone Y9C2UA
C;Genetics:
A;Gene: CESP:Y9C2UA.2
A;Map position: 2
A;Introns: 21/2
C;Superfamily: Caenorhabditis elegans hypothetical protein Y9C2UA.2

  Query Match             66.7%;  Score 4;  DB 2;  Length 103;
  Best Local Similarity   100.0%;  Pred. No. 1.3e+02;
  Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 RWHF 6
           ||||
Db      53 RWHF 56


RESULT 11
T44499
hypothetical protein [imported] - Thermomonospora fusca (fragment)
C;Species: Thermomonospora fusca
C;Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 21-Jan-2000
C;Accession: T44499
R;Irwin, D.I., Zhang, S.; Wilson, D.B.
submitted to the EMBL Data Library, April 1999
A;Description: Characterization of a Thermomonospora fusca family 48 exocellulase E6.
A;Reference number: Z22783
A;Accession: T44499
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-106 <IRW>
A;Cross-references: EMBL:AF144563; PIDN:AAD39950.1
A;Experimental source: strain YX

  Query Match             66.7%;  Score 4;  DB 2;  Length 106;
  Best Local Similarity   100.0%;  Pred. No. 1.3e+02;
  Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 WVRW 4
           ||||
Db      14 WVRW 17


RESULT 12
MNXR1B
sigma 1bNS protein - reovirus type 1 (strain Lang)
N;Alternate names: nonstructural protein
C;Species: reovirus type 1
```

```
A;Note: host Homo sapiens (man)
C;Date: 30-Jun-1988 #sequence_revision 30-Jun-1988 #text_change 16-Jul-1999
C;Accession: B26436
R;Munemitsu, S.M.; Atwater, J.A.; Samuel, C.E.
Biochem. Biophys. Res. Commun. 140, 528-534, 1986
A;Title: Biosynthesis of reovirus-specified polypeptides: nonvirion CPNS (sigma
peptide sigma-1s and the nonstructural polypeptide sigma 1bNS.
A;Reference number: A00129; MUID:87049799; PMID:2430868
A;Accession: B26436
A;Molecule type: mRNA
A;Residues: 1-119 <MUN>
A;Cross-references: GB:M14709; NID:g333744; PIDN:AAA47277.1; PID:g333746
A;Note: the authors translated the codon CAA for residue 5 as Glu
C;Genetics:
A;Map position: segment S1
C;Superfamily: reovirus sigma 1bNS protein
C;Keywords: nonstructural protein

  Query Match             66.7%;  Score 4;  DB 2;  Length 119;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+02;
  Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 WVRW 4
          ||||
Db      50 WVRW 53


RESULT 13
B34829
sigma 1bNS protein - reovirus type 1 (strain Lang)
N;Alternate names: nonstructural protein
C;Species: reovirus type 1
A;Note: host Homo sapiens (man)
C;Date: 08-Mar-1991 #sequence_revision 08-Mar-1991 #text_change 23-Sep-1999
C;Accession: B34829
R;Nibert, M.L.; Dermody, T.S.; Fields, B.N.
J. Virol. 64, 2976-2989, 1990
A;Title: Structure of the reovirus cell attachment protein: a model for the domain organ
A;Reference number: A34829; MUID:90244415; PMID:2335823
A;Accession: B34829
A;Status: preliminary
A;Molecule type: genomic RNA
A;Residues: 1-119 <NIB>
A;Cross-references: GB:M35963; NID:g333652; PIDN:AAA47243.1; PID:g333654
C;Genetics:
A;Map position: segment S1
C;Superfamily: reovirus sigma 1bNS protein

  Query Match             66.7%;  Score 4;  DB 2;  Length 119;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+02;
  Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 WVRW 4
          ||||
Db      50 WVRW 53


RESULT 14
T04366
probable peroxidase (EC 1.11.1.7), cationic - rice (fragment)
C;Species: Oryza sativa (rice)
C;Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 11-Jun-1999
C;Accession: T04366
R;Kim, C.Y.; Jeong, S.Y.; Choe, M.S.; Cheong, Y.H.; Han, C.D.; Cho, M.J.
submitted to the EMBL Data Library, March 1997
A;Description: Isolation and characterization of early inducible rice genes by a fungal
A;Reference number: Z15314
A;Accession: T04366
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-122 <KIM>
A;Cross-references: EMBL:U95217; NID:g2443458; PIDN:AAB71383.1; PID:g2443459
C;Genetics:
A;Gene: PRX1
C;Superfamily: peroxidase
C;Keywords: oxidoreductase

  Query Match             66.7%;  Score 4;  DB 2;  Length 122;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+02;
  Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 VRWH 5
          ||||
Db      50 VRWH 53


RESULT 15
B97400
hypothetical protein AGR_C_590 [imported] - Agrobacterium tumefaciens (strain C58, "Ver
C;Species: Agrobacterium tumefaciens
C;Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 11-Jan-2002
C;Accession: B97400
R;Goldner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman,
 A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.;
Science 294, 2323-2328, 2001
A;Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tume
A;Reference number: A97359; PMID:11743194
A;Accession: B97400
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-133 <KUR>
A;Cross-references: GB:AE009163; NID:g15161240; PIDN:AAK90159.1; PID:g15161241; GSPDB:GN00169
C;Genetics:
A;Gene: AGR_C_590
A;Map position: circular chromosome

  Query Match             66.7%;  Score 4;  DB 2;  Length 133;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+02;
  Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 VRWH 5
          ||||
Db     107 VRWH 110


Search completed: February 20, 2003, 11:09:39
Job time : 13 secs
```

GenCore version 5.1.3

OM protein - protein search, using sw model

Run on: February 20, 2003, 11:20:17 , Search time 11.3333 Seconds
(without alignments)
21.058 Million cell updates/sec

Title: US-09-912-414-2
Perfect score: 6
Sequence: 1 WVRWHF 6

Scoring table: OLIGODX
Gapop 60.0 , Gapext 60.0

Searched: 112892 seqs, 41476328 residues

Word size : 0

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No	Score	% Query Match	Length	DB	ID	Description
1	5	83.3	372	1	GP45_HUMAN	Q9y5y3 homo sapien
2	5	83.3	373	1	GP45_MOUSE	Q9eqq4 mus musculu
3	5	83.3	799	1	AFSK_STRCO	P54741 streptomyce
4	5	83.3	807	1	AFSK_STRGR	P54742 streptomyce
5	4	66.7	95	1	YPDA_BACST	P21878 bacillus st
6	4	66.7	119	1	VSIS_REOVL	P07938 reovirus (t
7	4	66.7	149	1	GLB1_MOPMP	P21197 mordacia mo
8	4	66.7	149	1	GLB1_PETMA	P09967 petromyzon
9	4	66.7	149	1	GLB2_MOPMP	P21198 mordacia mo
10	4	66.7	149	1	GLB2_PETMA	Q9i9i3 petromyzon
11	4	66.7	149	1	GLB3_MOPMP	P21199 mordacia mo
12	4	66.7	149	1	GLB3_PETMA	P09968 petromyzon
13	4	66.7	149	1	GLB5_PETMA	P02208 petromyzon
14	4	66.7	149	1	GLB_LAMFL	P02207 lampetra fl
15	4	66.7	170	1	CBP4_YEAST	P37267 saccharomyc
16	4	66.7	195	1	TNF7_MOUSE	O55237 mus musculu
17	4	66.7	197	1	DSBE_XYLFA	Q9pan4 xylella fas
18	4	66.7	208	1	YQED_BACSU	P54449 bacillus su
19	4	66.7	215	1	UR2P_BOVIN	P49220 bos taurus
20	4	66.7	223	1	PGC2_HUMAN	O15173 homo sapien
21	4	66.7	244	1	YU68_CAEEL	P52057 caenorhabdi
22	4	66.7	286	1	PURU_CORSP	Q46339 corynebacte
23	4	66.7	295	1	MOPA_PSEPU	Q02198 pseudomonas
24	4	66.7	298	1	YIHV_ECOLI	P32143 escherichia
25	4	66.7	310	1	DNJL_MYCGE	P47248 mycoplasma
26	4	66.7	322	1	ASTB_ECOLI	P76215 escherichia
27	4	66.7	325	1	UL76_HCMVA	P16725 human cytom
28	4	66.7	338	1	PDXA_RALSO	P58714 ralstonia s
29	4	66.7	389	1	YM67_MYCTU	O06625 mycobacteri
30	4	66.7	403	1	Y4HM_RHISN	P55480 rhizobium s
31	4	66.7	418	1	PHAA_BACHD	Q9kcl9 bacillus ha
32	4	66.7	422	1	TERI_PRSPF	P54308 bacteriopha
33	4	66.7	436	1	SOC4_MOUSE	Q61245 mus musculu
34	4	66.7	440	1	SOC4_HUMAN	Q8wxh5 homo sapien
35	4	66.7	450	1	Y0E9_PSEAE	Q9hxy3 pseudomonas
36	4	66.7	462	1	SYTM_YEAST	P07246 saccharomyc
37	4	66.7	472	1	PPB_ESCFE	P21948 escherichia
38	4	66.7	475	1	ETS6_DROME	P29776 drosophila
39	4	66.7	486	1	YDFI_ECOLI	P77260 escherichia
40	4	66.7	507	1	TRA6_PSEAE	Q57541 pseudomonas
41	4	66.7	508	1	TRAT_CHEHE	Q46087 cheliclast
42	4	66.7	511	1	FAST_MOUSE	Q9jix9 mus musculu
43	4	66.7	549	1	FAST_HUMAN	Q14296 homo sapien
44	4	66.7	583	1	STS_HUMAN	P08842 homo sapien
45	4	66.7	611	1	SNF1_CANGA	Q00372 candida gla

ALIGNMENTS

RESULT 1
GP45_HUMAN
ID GP45_HUMAN STANDARD; PRT; 372 AA.
AC Q9Y5Y3;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Probable G-protein-coupled receptor GPR45 (PSP24-alpha) (PSP24-1).
GN GPR45.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99156852; PubMed=10026181;
RA Marchese A., Sawzdargo M., Nguyen T., Cheng R., Heng H.H., Nowak T.,
RA Im D.-S., Lynch K.R., George S.R., O'Dowd B.F.;
RT "Discovery of three novel orphan G-protein-coupled receptors.";
RL Genomics 56:12-21(1999).
CC -!- FUNCTION: Orphan receptor. May play a role in brain function.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC -!- TISSUE SPECIFICITY: Expressed in brain; detected in the basal
CC forebrain, frontal cortex, and caudate, but not in thalamus,
CC hippocampus, or putamen.
CC -!- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
CC ---------- ... ----------
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC ---------- ... ----------
DR EMBL; AF119066, AAD21056.1; -.
DR GeneCew, HGNC:4503, GPR45.
DR MIM; 604838; -.
DR InterPro; IPR000276; GPCR_Rhodpsn.
DR Pfam; PF00001; 7tm_1; 1.
DR PRINTS; PR00237; GPCRPHODOPSN.
DR PROSITE; PS00237; G_PROTEIN_RECEP_F1_1; FALSE_NEG.
DR PROSITE; PS50262; G_PROTEIN_RECEP_F1_2; 1.
KW G-protein coupled receptor; Transmembrane; Glycoprotein.
FT DOMAIN 1 38 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 39 59 1 (POTENTIAL).
FT DOMAIN 60 75 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 76 96 2 (POTENTIAL).
FT DOMAIN 97 109 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 110 130 3 (POTENTIAL).
FT DOMAIN 131 149 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 150 170 4 (POTENTIAL).
FT DOMAIN 171 198 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 199 219 5 (POTENTIAL).
FT DOMAIN 220 268 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 269 289 6 (POTENTIAL).

```
FT   DOMAIN       299    305       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     306    326       1 (POTENTIAL).
FT   DOMAIN       327    372       CYTOPLASMIC (POTENTIAL).
FT   CARBOHYD       4      4       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD      17     17       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD      20     20       N-LINKED (GLCNAC...) (POTENTIAL).
SQ   SEQUENCE   372 AA;  42000 MW;  B4B4EC7AEDE20199 CRC64;

  Query Match             83.3%;  Score 5;  DB 1;  Length 372;
  Best Local Similarity   100.0%;  Pred. No. 4.4;
  Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 VRWHF 6
             |||||
Db      98 VRWHF 102


RESULT 2
GP45_MOUSE
ID   GP45_MOUSE     STANDARD;      PRT;   373 AA.
AC   Q9EQQ4;
DT   15-JUN-2000 (Rel. 41, Created)
DT   15-JUN-2001 (Rel. 41, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Probable G-protein-coupled receptor GPR45 (PSP24-alpha) (PSP24-1).
GN   GPR45.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20483643; PubMed-11027574;
RA   Kawasawa Y., Kume K., Nakade S., Haga H., Izumi T., Shimizu T.,
RT   "Brain-specific expression of novel G-protein-coupled receptors, with
RT   homologies to Xenopus PSP24 and human GPR45.",
RL   Biochem. Biophys. Res. Commun. 276:1-6 (2000).
CC   -!- FUNCTION: Orphan receptor. May play a role in brain function.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- TISSUE SPECIFICITY: Brain-specific.
CC   -!- SIMILARITY: BELONGS TO FAMILY 1 OF G PROTEIN COUPLED RECEPTORS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF139642; AAG40176.1; -.
DR   MGD; MGI:2135682; Gpr45.
DR   InterPro; IPR000276; GPCR_Rhodpsn.
DR   Pfam; PF00001; 7tm_1; 1.
DR   PRINTS; PR00237; GPCRRHODOPSN.
DR   PROSITE; PS00237; G_PROTEIN_RECEP_F1_1; FALSE_NEG.
DR   PROSITE; PS50262; G_PROTEIN_RECEP_F1_2; 1.
KW   G-protein coupled receptor; Transmembrane; Glycoprotein.
FT   DOMAIN        1     38       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     39     59       1 (POTENTIAL)
FT   DOMAIN       60     75       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     76     96       2 (POTENTIAL).
FT   DOMAIN       97    109       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    110    130       3 (POTENTIAL).
FT   DOMAIN      131    147       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    148    170       4 (POTENTIAL).
FT   DOMAIN      171    198       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    199    219       5 (POTENTIAL).
FT   DOMAIN      220    269       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    270    290       6 (POTENTIAL).
FT   DOMAIN      291    306       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    307    327       7 (POTENTIAL).
FT   DOMAIN      328    372       CYTOPLASMIC (POTENTIAL).
FT   CARBOHYD     17     17       N-LINKED (GLCNAC...) (POTENTIAL).
SQ   SEQUENCE   373 AA;  42308 MW;  693BE1CB3673FBAD CRC64;

  Query Match             83.3%;  Score 5;  DB 1;  Length 373;
  Best Local Similarity   100.0%;  Pred. No. 4.4;
  Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 VRWHF 6
             |||||
Db      98 VRWHF 102


RESULT 3
AFSK_STRCO
ID   AFSK_STRCO     STANDARD;      PRT;   799 AA.
AC   P54741; Q9L002; Q9F365;
DT   01-OCT-1996 (Rel. 34, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Serine/threonine protein kinase afsK (EC 2.7.1.-).
GN   AFSK OR SCO4423 OR SC6F11.21 OR SCD6.01.
OS   Streptomyces coelicolor.
OC   Bacteria; Actinobacteria, Actinobacteridae; Actinomycetales;
OC   Actinomycetineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=1902;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2);
RX   MEDLINE=94241108; PubMed-8063064;
RA   Matsumoto A., Hong S.K., Ishizuka H., Horinouchi S., Beppu T.;
RT   "Phosphorylation of the AfsR protein involved in secondary metabolism
RT   in Streptomyces species by a eukaryotic type protein kinase.";
RL   Gene 146:47-56(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2);
RX   MEDLINE=98126767; PubMed-9461567;
RA   Ueda K., Umeyama T., Beppu T., Horinouchi S.;
RT   "The aerial mycelium-deficient phenotype of Streptomyces griseus
RT   resulting from A factor deficiency is suppressed by a Ser/Thr kinase
RT   of S. coelicolor A3(2).";
RL   Gene 169:91-96(1996).
RN   [3]
RP   IDENTITY TO 233-240.
RA   Matsumoto A., Hong S., Ishizuka H., Horinouchi S., Beppu T.,
RA   Umeyama T.;
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2) / M145;
RA   MEDLINE=21130133; PubMed-11600533;
RA   Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
RA   Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
RA   Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA   Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA   Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neil S.,
RA   Rabbinowitsch E., Rajandream M.-A., Rutherford K., Rutter S.,
RA   Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA   Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
RA   Hopwood D.A.;
RT   "Complete genome sequence of the model actinomycete Streptomyces
RT   coelicolor A3(2).";
RL   Nature 417:141-147(2002).
CC   -!- FUNCTION: INVOLVED IN THE REGULATION OF SECONDARY METABOLISM BY
CC       PHOSPHORYLATING, ON BOTH SER AND THR, THE AFSR GLOBAL REGULATORY
CC       PROTEIN INVOLVED IN THE CONTROL OF SECONDARY METABOLISM.
CC   -!- PTM: AUTOPHOSPHORYLATED AT THR AND SER RESIDUES.
CC   -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
```

```
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; D45382; BAA08220.2; -.
DR   EMBL; AL392178; CAC08432.1; -.
DR   EMBL; AL353815; CAB88430.1; -.
DR   PhosSite; P54741; -.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR002290; Ser_thr_pkinase.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   Pfam; PF00069; pkinase; 1.
DR   ProDom; PD000001; Euk_pkinase; 1.
DR   SMART; SM00220; S_TKc; 1.
DR   SMART; SM00219; TyrKc; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW   Transferase; Serine/threonine-protein kinase; ATP binding;
KW   Phosphorylation; Complete proteome.
FT   DOMAIN       16    271       PROTEIN KINASE.
FT   NP_BIND      22     30       ATP (BY SIMILARITY).
FT   BINDING      44     44       ATP (BY SIMILARITY).
FT   ACT_SITE    138    138       BY SIMILARITY.
SQ   SEQUENCE   799 AA;  88787 MW;  4FE9FED41B9FB5BE CRC64;

  Query Match             83.3%;  Score 5;  DB 1;  Length 799;
  Best Local Similarity   100.0%;  Pred. No. 8.6;
  Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 VRWHF 6
           |||||
Db     628 VRWHF 632


RESULT 4
AFSK_STRGR
ID   AFSK_STRGR      STANDARD;      PRT;   807 AA.
AC   P54742;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   Serine/threonine protein kinase afsK (EC 2.7.1 -)
GN   AFSK.
OS   Streptomyces griseus.
OC   Bacteria; Actinobacteria, Actinobacteria (class); Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces
OX   NCBI_TaxID=1911;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=96186909; PubMed=8635767;
RA   Ueda K., Umeyama T., Beppu T., Horinouchi S.;
RT   "The aerial mycelium-defective phenotype of Streptomyces griseus
RT   resulting from A-factor deficiency is suppressed by a Ser/Thr kinase
RT   of S. coelicolor A3(2).";
RL   Gene 169:91-95(1996).
CC   -!- FUNCTION: INVOLVED IN THE REGULATION OF SECONDARY METABOLISM BY
CC       PHOSPHORYLATING, ON BOTH SER AND THR, THE AFSR GLOBAL REGULATORY
CC       PROTEIN INVOLVED IN THE CONTROL OF SECONDARY METABOLISM (BY
CC       SIMILARITY).
CC   -!- PTM: AUTOPHOSPHORYLATED AT THR AND SER RESIDUES (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright.  It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; D45246; BAA08203.1; -.
DR   PhosSite; P54742; -.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR002290; Ser_thr_pkinase.
DR   Pfam; PF00069; pkinase; 1.
DR   ProDom; PD000001; Euk_pkinase; 1.
DR   PROSITE, PS00107, PROTEIN_KINASE_ATP; 1.
DR   PROSITE, PS00108, PROTEIN_KINASE_ST; 1.
DR   PROSITE, PS50011, PROTEIN_KINASE_DOM; 1.
KW   Transferase; Serine/threonine-protein kinase; ATP-binding;
KW   Phosphorylation.
FT   DOMAIN       16    272       PROTEIN KINASE.
FT   NP_BIND      22     30       ATP (BY SIMILARITY).
FT   BINDING      44     44       ATP (BY SIMILARITY).
FT   ACT_SITE    138    138       BY SIMILARITY.
SQ   SEQUENCE   807 AA,  88231 MW,  [illegible] CRC64;

  Query Match             83.3%;  Score 5;  DB 1;  Length 807;
  Best Local Similarity   100.0%;  Pred. No. 8.7;
  Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 VRWHF 6
           |||||
Db     635 VRWHF 639


RESULT 5
YPDA_BACST
ID   YPDA_BACST      STANDARD;      PRT;    95 AA.
AC   P21878;
DT   01-MAY-1991 (Rel. 18, Created)
DT   01-MAY-1991 (Rel. 18, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hypothetical protein in pdhA 5'region (ORF1) (Fragment).
OS   Bacillus stearothermophilus.
OC   Bacteria; Firmicutes; Bacillales; Geobacillus.
OX   NCBI_TaxID=1422;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NCA 1503;
RX   MEDLINE=90345939; PubMed=2200674;
RA   Hawkins C.F., Borges A., Perham R.N.;
RT   "Cloning and sequence analysis of the genes encoding the alpha and
RT   beta subunits of the E1 component of the pyruvate dehydrogenase
RT   multienzyme complex of Bacillus stearothermophilus.";
RL   Eur. J. Biochem. 191:337-346(1990).
CC   -!- SIMILARITY: BELONGS TO THE YOF/YPHA/YIDA/YIND (E.COLI) / YOFE/YXEH
CC       (B.SUBTILIS) FAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; X53560; CAA37626.1; -.
DR   PIR; S10796; S10796.
DR   InterPro, IPR001454, Hignase/hydrlase.
DR   InterPro; IPR000150; Hypothet_cof.
DR   Pfam; PF00702; Hydrolase; 1.
DR   PROSITE, PS01228, COF_1, PARTIAL.
DR   PROSITE, PS01229, COF_2; 1.
KW   Hypothetical protein.
FT   NON_TER       1      1
SQ   SEQUENCE   95 AA;  10441 MW;  811E9487321A74E3 CRC64;

  Query Match             66.7%;  Score 4;  DB 1;  Length 95;
  Best Local Similarity   100.0%;  Pred. No. 27;
  Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 VRWH 5
           ||||
Db       5 VRWH 8
```

```
RESULT 6
VGIC_REOVL
ID   VGIC_REOVL     STANDARD;      PRT;   119 AA.
AC   P07938;
DT   01-AUG-1988 (Rel. 08, Created)
DT   01-AUG-1988 (Rel. 08, Last sequence update)
DT   01-FEB-1996 (Rel. 33, Last annotation update)
DE   Sigma 1s protein.
OS   Reovirus (type 1 / strain Lang).
OC   Viruses; dsRNA viruses; Reoviridae; Orthoreovirus.
OX   NCBI_TaxID=10884;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=87048799; PubMed-2433568;
RA   Muhemitsu S.M., Atwater J.A., Samuel C.E.;
RT   "Biosynthesis of reovirus-specified polypeptides. Molecular cDNA
RT   cloning and nucleotide sequence of the termini of reovirus type 1 Lang
RT   strain bicistronic S1 mRNA which encodes the minor capsid polypeptide
RT   sigma 1s and the nonstructural polypeptide sigma 1NS.";
RL   Biochem. Biophys. Res. Commun. 140:508-514(1986).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20244415; PubMed-2335933;
RA   Nibert M.L., Dermody T.S., Fields B.M.;
RT   "Structure of the reovirus cell-attachment protein: a model for the
RT   domain organization of sigma 1.";
RL   J. Virol. 64:2976-2989(1990).
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=85115945; PubMed-3885345;
RA   Cashdollar L.W., Chmelo R.A., Wiener J.R., Joklik W.K.;
RT   "Sequence of the S1 genes of the three serotypes of reovirus.";
RL   Proc. Natl. Acad. Sci. U.S.A. 82:24-28(1985).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collab-
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announ
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M24773; AAA47207.1; -.
DR   EMBL; M35963; AAA47243.1; -.
DR   EMBL; M10260; AAA66878.1; -.
DR   PIR; A29429; MNRA1S.
DR   InterPro; IPR003478; Sigma_1s.
DR   Pfam; PF02454; Sigma_1s; 1.
FT   CONFLICT     71     71       V -> L (IN REF. 2).
SQ   SEQUENCE   119 AA;  13666 MW;  DABBEB1D5E8C1D1 CRC64;

  Query Match             66.7%;  Score 4;  DB 1;  Length 119;
  Best Local Similarity   100.0%;  Pred. No. 40;
  Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 WVRW 4
           |||
Db      50 WVRW 53


RESULT 7
GLB1_MORMR
ID   GLB1_MORMR     STANDARD;      PRT;   149 AA.
AC   P21197;
DT   01-MAY-1991 (Rel. 18, Created)
DT   01-MAY-1991 (Rel. 18, Last sequence update)
DT   15-JUN-2001 (Rel. 41, Last annotation update)
DE   Globin I.
OS   Mordacia mordax (Southern hemisphere lamprey).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Hyperoartia;
OC   Petromyzontiformes; Petromyzontidae; Mordacia.
OX   NCBI_TaxID=7755;
RN   [1]
RP   SEQUENCE.
RX   MEDLINE=91248417; PubMed-2039605;
RA   Hendriades I., Vidal Y., Bodewald K., Braunitzer G., Nouzil E.;
RT   "The primary structure of the hemoglobins of a southern hemisphere
RT   lamprey (Mordacia mordax, Cyclostomata).";
RL   Biol. Chem. Hoppe-Seyler 372:49-56(1991).
CC   -!- SUBUNIT: Monomer.
CC   -!- SIMILARITY: BELONGS TO THE GLOBIN FAMILY.
DR   PIR; S13458; S13458.
DR   HSSF; PS2208; 2LHB.
DR   InterPro; IPR000971; Globin.
DR   Pfam; PF00042; globin; 1.
DR   PROSITE; PS01033; GLOBIN; 1.
KW   Heme; Oxygen transport; Transport.
SQ   SEQUENCE   149 AA;  16498 MW;  A1A7B8B54F55F58E CRC64;

  Query Match             66.7%;  Score 4;  DB 1;  Length 149;
  Best Local Similarity   100.0%;  Pred. No. 40;
  Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 VRWH 5
           |||
Db      70 VRWH 73


RESULT 8
GLB1_PETMA
ID   GLB1_PETMA     STANDARD;      PRT;   149 AA.
AC   P09967;
DT   01-MAR-1989 (Rel. 10, Created)
DT   01-MAR-1989 (Rel. 10, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Globin 1.
OS   Petromyzon marinus (Sea lamprey).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Hyperoartia;
OC   Petromyzontiformes; Petromyzontidae; Petromyzon.
OX   NCBI_TaxID=7757;
RN   [1]
RP   SEQUENCE.
RX   MEDLINE=87184913; PubMed-3566915;
RA   Hendriades I., Rudewald M., Allard M., Nouzil E., Braunitzer G.;
RT   "Primary structure of the minor haemoglobins from the sea lamprey
RT   (Petromyzon marinus, Cyclostomata).";
RL   Biol. Chem. Hoppe-Seyler 368:145-154(1987).
CC   -!- SUBUNIT: Monomer.
CC   -!- MISCELLANEOUS: THIS IS ONE OF THE MINOR GLOBIN COMPONENT OF SEA
CC       LAMPREY.
CC   -!- SIMILARITY: BELONGS TO THE GLOBIN FAMILY.
DR   PIR; A26042; A26042.
DR   HSSF; PS2208; 2LHB.
DR   InterPro; IPR000971; Globin.
DR   Pfam; PF00042; globin; 1.
DR   PROSITE; PS01033; GLOBIN; 1.
KW   Heme; Oxygen transport; Transport.
SQ   SEQUENCE   149 AA;  16498 MW;  CBCC39B31C4CF846 CRC64;

  Query Match             66.7%;  Score 4;  DB 1;  Length 149;
  Best Local Similarity   100.0%;  Pred. No. 40;
  Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 VRWH 5
           |||
Db      70 VRWH 73


RESULT 9
GLB2_MORMR
ID   GLB2_MORMR     STANDARD;      PRT;   149 AA.
AC   P21198;
DT   01-MAY-1991 (Rel. 18, Created)
DT   01-MAY-1991 (Rel. 18, Last sequence update)
```

```
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Globin II.
OS   Mordacia mordax (Southern hemisphere lamprey).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Hyperoartia;
OC   Petromyzontiformes; Petromyzontidae; Mordacia.
OX   NCBI_TaxID=7755;
RN   [1]
RP   SEQUENCE.
RX   MEDLINE=91248417; PubMed=2039605;
RA   Hombrados I., Vidal Y., Rodewald K., Braunitzer G., Neuzil E.;
RT   "The primary structure of the hemoglobins of a southern hemisphere
RT   lamprey (Mordacia mordax, Cyclostomata).";
RL   Biol. Chem. Hoppe-Seyler 372:49-56(1991).
CC   -!- SUBUNIT: Monomer.
CC   -!- SIMILARITY: BELONGS TO THE GLOBIN FAMILY.
DR   PIR; S13459; S13459.
DR   HSSP; P02208; 2LHB.
DR   InterPro; IPR000971; Globin.
DR   Pfam; PF00042; globin; 1.
DR   PROSITE; PS01033; GLOBIN; 1.
KW   Heme; Oxygen transport; Transport.
SQ   SEQUENCE   149 AA;  16498 MW;  [illegible] CRC64;

  Query Match             66.7%;  Score 4;  DB 1;  Length 149;
  Best Local Similarity   100.0%;  Pred. No. 40;
  Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 VRWH 5
           ||||
Db      70 VRWH 73


RESULT 10
GLB2_PETMA
ID   GLB2_PETMA     STANDARD;      PRT;   149 AA.
AC   Q9I9I3;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Globin II (Hemoglobin PMII').
OS   Petromyzon marinus (Sea lamprey).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Hyperoartia;
OC   Petromyzontiformes; Petromyzontidae; Petromyzon.
OX   NCBI_TaxID=7757;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20250908; PubMed=10788466;
RA   Qiu Y., Maillett D.H., Knapp J., Olson J.S., Riggs A.F.;
RT   "Lamprey hemoglobin. Structural basis of the Bohr effect.";
RL   J. Biol. Chem. 275:13517-13528(2000).
CC   -!- SIMILARITY: BELONGS TO THE GLOBIN FAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AF248645; AAF67186.1;  -.
DR   HSSP; P02208; 2LHB.
DR   InterPro; IPR000971; Globin.
DR   InterPro; IPR002335; Myoglobin.
DR   Pfam; PF00042; globin; 1.
DR   PRINTS; PR00613; MYOGLOBIN.
DR   PROSITE; PS01033; GLOBIN; 1.
KW   Heme; Oxygen transport; Transport.
FT   INIT_MET      0      0
SQ   SEQUENCE   149 AA;  16311 MW;  [illegible] CRC64;

  Query Match             66.7%;  Score 4;  DB 1;  Length 149;
  Best Local Similarity   100.0%;  Pred. No. 40;
  Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 VRWH 5
           ||||
Db      70 VRWH 73


RESULT 11
GLB3_MORMR
ID   GLB3_MORMR     STANDARD;      PRT;   149 AA.
AC   P21199;
DT   01-MAY-1991 (Rel. 18, Created)
DT   01-MAY-1991 (Rel. 18, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Globin III.
OS   Mordacia mordax (Southern hemisphere lamprey).
OC   Eukaryota, Metazoa, Chordata, Craniata, Vertebrata, Hyperoartia;
OC   Petromyzontiformes; Petromyzontidae, Mordacia.
OX   NCBI_TaxID=7755;
RN   [1]
RP   SEQUENCE.
RX   MEDLINE=91248417; PubMed=2039605;
RA   Hombrados I., Vidal Y., Rodewald K., Braunitzer G., Neuzil E.;
RT   "The primary structure of the hemoglobins of a southern hemisphere
RT   lamprey (Mordacia mordax, Cyclostomata).";
RL   Biol. Chem. Hoppe-Seyler 372:49-56(1991).
CC   -!- SUBUNIT: Monomer.
CC   -!- SIMILARITY: BELONGS TO THE GLOBIN FAMILY
DR   PIR; S13460; S13460.
DR   HSSP; P02208; 2LHB.
DR   InterPro; IPR000971; Globin.
DR   Pfam; PF00042; globin; 1.
DR   PROSITE; PS01033; GLOBIN; 1.
KW   Heme; Oxygen transport; Transport.
SQ   SEQUENCE   149 AA;  16598 MW;  [illegible] CRC64;

  Query Match             66.7%;  Score 4;  DB 1;  Length 149;
  Best Local Similarity   100.0%;  Pred. No. 40;
  Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 VRWH 5
           ||||
Db      70 VRWH 73


RESULT 12
GLB3_PETMA
ID   GLB3_PETMA     STANDARD;      PRT;   149 AA.
AC   P09968;
DT   01-MAR-1989 (Rel. 10, Created)
DT   01-MAR-1989 (Rel. 10, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Globin III.
OS   Petromyzon marinus (Sea lamprey).
OC   Eukaryota, Metazoa, Chordata, Craniata, Vertebrata; Hyperoartia;
OC   Petromyzontiformes; Petromyzontidae; Petromyzon.
OX   NCBI_TaxID=7757;
RN   [1]
RP   SEQUENCE.
RX   MEDLINE=87184913; PubMed=3566915;
RA   Hombrados I., Rodewald K., Allard M., Neuzil E., Braunitzer G.;
RT   "Primary structure of the minor haemoglobins from the sea lamprey
RT   (Petromyzon marinus, Cyclostomata).";
RL   Biol. Chem. Hoppe-Seyler 368:145-154(1987).
CC   -!- SUBUNIT: Monomer.
CC   -!- MISCELLANEOUS: THIS IS ONE OF THE MINOR GLOBIN COMPONENT OF SEA
CC       LAMPREY.
CC   -!- SIMILARITY: BELONGS TO THE GLOBIN FAMILY.
DR   PIR; B26042; B26042.
DR   HSSP; P02208; 2LHB.
DR   InterPro; IPR000971; Globin.
DR   Pfam; PF00042; globin; 1.
DR   PROSITE; PS01033; GLOBIN; 1.
```

```
KW   Heme; Oxygen transport; Transport.
SQ   SEQUENCE   149 AA;  16811 MW;  A9A33B4412C917AE CRC64;

  Query Match           66.7%;  Score 4;  DB 1;  Length 149;
  Best Local Similarity  100.0%;  Pred. No. 40;
  Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 VRWH 5
           ||||
Db      70 VRWH 73


RESULT 13
GLB5_PETMA
ID   GLB5_PETMA     STANDARD;      PRT;   149 AA.
AC   P02208;
DT   21-JUL-1986 (Rel. 01, Created)
DT   21-JUL-1986 (Rel. 01, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Globin V.
OS   Petromyzon marinus (Sea lamprey).
OC   Eukaryota, Metazoa, Chordata, Craniata, Vertebrata, Hyperartia,
OC   Petromyzontiformes, Petromyzontidae, Petromyzon.
OX   NCBI_TaxID=7757;
RN   [1]
RP   SEQUENCE.
RX   MEDLINE=83257424; PubMed=6409159;
RA   Hendrickson T., Redbewald F., Neuzil E., Braunitzer G.,
RT   "Haemoglobins LX. Primary structure of the major haemoglobin of the
RT   sea lamprey Petromyzon marinus (var. Garonne, Loire).";
RL   Biochimie 65:247-257(1983).
RN   [2]
RP   SEQUENCE.
RX   MEDLINE=71051151; PubMed=5484471;
RA   Li S.L., Riggs A.;
RT   "The amino acid sequence of hemoglobin V from the lamprey, Petromyzon
RT   marinus.";
RL   J. Biol. Chem. 245:6140-6169(1970).
RN   [3]
RP   X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).
RX   MEDLINE=73143009; PubMed=4692995;
RA   Hendrickson W.A., Love W.E., Karle J.,
RT   "Crystal structure analysis of sea lamprey hemoglobin at 2 A
RT   resolution.";
RL   J. Mol. Biol. 74:331-361(1973).
RN   [4]
RP   X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).
RX   MEDLINE=85203179; PubMed=4024496;
RA   Honzatko R.B., Hendrickson W.A., Love W.E.;
RT   "Refinement of a molecular model for lamprey hemoglobin from
RT   Petromyzon marinus",
RL   J. Mol. Biol. 184:147-164(1985).
RN   [5]
RP   X-RAY CRYSTALLOGRAPHY (2.7 ANGSTROMS).
RX   MEDLINE=99306034; PubMed=10378271;
RA   Heaslet H.A., Royer W.E. Jr.,
RT   "The 2.7-A crystal structure of deoxygenated hemoglobin from the sea
RT   lamprey, Petromyzon marinus: structural basis for a lowered oxygen
RT   affinity and Bohr effect.",
RL   Structure 7:517-526(1999).
CC   -!- FUNCTION: MINIMUM AT HIGH OXYGEN TENSION AND HIGH PH AND DIMERIZ
CC       AT LOW OXYGEN TENSION AND LOWER PH.
CC   -!- MISCELLANEOUS: GLOBIN V IS THE MAJOR COMPONENT OF THE 4 GLOBINS
CC       FOUND IN THE SEA LAMPREY.
CC   -!- SIMILARITY: BELONGS TO THE GLOBIN FAMILY.
DR   PIR; A02529; GGLMS.
DR   PDB; 2LHB; 15-APR-91.
DR   PDB; 3LHB; 20-JAN-99.
DR   InterPro; IPR000971; Globin.
DR   Pfam; PF00042; globin; 1.
DR   PROSITE; PS01033; GLOBIN; 1.
KW   Heme; Oxygen transport; Transport; 3D-structure.
FT   VARIANT      29     29       T -> N.
FT   CONFLICT     29     29       S -> SS (IN REF. 3).
FT   CONFLICT     98     99       MISSING (IN REF. 2 AND 3).
FT   CONFLICT    100    100       D -> N (IN REF. 2).
FT   CONFLICT    138    138       M -> R (IN REF. 2 AND 3).
FT   CONFLICT    139    139       MISSING (IN REF. 2).
FT   HELIX        13     28
FT   TURN         29     29
FT   HELIX        30     44
FT   HELIX        46     51
FT   HELIX        53     55
FT   TURN         56     57
FT   HELIX        61     65
FT   TURN         66     66
FT   HELIX        68     86
FT   TURN         87     89
FT   HELIX        91    107
FT   TURN        108    108
FT   HELIX       113    128
FT   TURN        129    130
FT   HELIX       132    146
FT   TURN        147    148
SQ   SEQUENCE   149 AA;  16811 MW;  7F1331FB3CCA2EE3 CRC64;

  Query Match           66.7%;  Score 4;  DB 1;  Length 149;
  Best Local Similarity  100.0%;  Pred. No. 40;
  Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 VRWH 5
           ||||
Db      70 VRWH 73


RESULT 14
GLB_LAMFL
ID   GLB_LAMFL      STANDARD;      PRT;   149 AA.
AC   P02207;
DT   21-JUL-1986 (Rel. 01, Created)
DT   21-JUL-1986 (Rel. 01, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Globin.
OS   Lampetra fluviatilis (River lamprey).
OC   Eukaryota, Metazoa, Chordata, Craniata, Vertebrata, Hyperartia;
OC   Petromyzontiformes, Petromyzontidae; Lampetra.
OX   NCBI_TaxID=7748;
RN   [1]
RP   SEQUENCE.
RX   MEDLINE=82114121; PubMed=118914;
RA   Zelenik M., Radliff V., Braunitzer G.;
RT   "Hemoglobins, XXX. The amino acid sequence of the monomeric
RT   hemoglobin from Lampetra fluviatilis.";
RL   Hoppe Seyler's Z. Physiol. Chem. 361:1879-1894(1979).
CC   -!- SUBUNIT: Monomer.
CC   -!- SIMILARITY: BELONGS TO THE GLOBIN FAMILY.
DR   PIR; A02529; GGLMF.
DR   HSSP; P02208; 2LHB.
DR   InterPro; IPR000971; Globin.
DR   Pfam; PF00042; globin; 1.
DR   PROSITE; PS01033; GLOBIN; 1.
KW   Heme; Oxygen transport; Transport; Phosphorylation.
FT   MOD_RES       1      1       FORMYLATION (PARTIAL).
SQ   SEQUENCE   149 AA;  16813 MW;  B01FFAFFA3A2CC3F CRC64;

  Query Match           66.7%;  Score 4;  DB 1;  Length 149;
  Best Local Similarity  100.0%;  Pred. No. 40;
  Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 VRWH 5
           ||||
Db      70 VRWH 73


RESULT 15
CBP4_YEAST
```

```
ID   CBP4_YEAST     STANDARD;      PRT;   170 AA.
AC   P37267;
DT   01-OCT-1994 (Rel. 30, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   CBP4 protein, mitochondrial precursor.
GN   CBP4 OR YGR174C.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota, Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales, Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94342301; PubMed=8063753;
RA   Crivellone M.D.;
RT   "Characterization of CBP4, a new gene essential for the expression of
RT   ubiquinol-cytochrome c reductase in Saccharomyces cerevisiae.";
RL   J. Biol. Chem. 269:21284-21292(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Hebling U., Hofmann B., Delius H.;
RL   Submitted (MAY-1996) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: ESSENTIAL FOR THE ASSEMBLY AND/OR STABILITY OF
CC       UBIQUINOL-CYTOCHROME C REDUCTASE. IT HAS A DIRECT EFFECT ON THE
CC       CORRECT ASSEMBLY OF THE RIESKE PROTEIN, CORE 4, CORE 5 AND
CC       APOCYTOCHROME B; IT MAY EITHER BE INVOLVED IN POST-TRANSLATIONAL
CC       MODIFICATION OF THE SUBUNITS OR IN THE ASSEMBLY OF THE ENZYME.
CC   -!- SUBCELLULAR LOCATION: MITOCHONDRIAL, ASSOCIATED WITH THE INNER
CC       MEMBRANE.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U10700; AAA61566.1; -.
DR   EMBL; Z72959; CAA97200.1; -.
DR   PIR; A53928; A53928.
DR   SGD; S0003406; CBP4.
KW   Mitochondrion; Transit peptide.
FT   TRANSIT       1      ?       MITOCHONDRION (POTENTIAL).
FT   CHAIN         ?    170       CBP4 PROTEIN.
FT   CONFLICT     65     65       S -> F (IN REF. 1).
SQ   SEQUENCE   170 AA;  20219 MW;  D88E02EADF0B366B CRC64;

  Query Match             66.7%;  Score 4;  DB 1;  Length 170;
  Best Local Similarity   100.0%;  Pred. No. 44;
  Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 WVRW 4
           ||||
Db      29 WVRW 32


Search completed: February 20, 2003, 11:24:56
Job time : 12.3333 secs
```

GenCore version 5.1.3

OM protein - protein search, using sw model

Run on: February 20, 2003, 11:23:27 ; Search Time 47 Seconds
(without alignments)
26.876 Million cell updates/sec

Title: US-09-912-414-2
Perfect score: 6
Sequence: 1 WVRWHF 6

Scoring table: OLIGODX
Gapop 60.0 , Gapext 60.0

Searched: 671580 seqs, 200047115 residues

Word size: 0

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : SPTREMBL_21:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution

SUMMARIES

Result No	Score	% Query Match	Length	DB	ID	Description
1	5	83.3	301	2	O52770	O52770 pseudomonas
2	5	83.3	301	16	Q9I1E1	Q9i1e1 pseudomonas
3	5	83.3	308	6	Q8WNV6	Q8wnv6 capra hircu
4	5	83.3	366	16	Q9FC71	Q9fc71 streptomyce
5	5	83.3	373	10	Q9M9G2	Q9m9g2 arabidopsis
6	5	83.3	376	4	O43898	O43898 homo sapien
7	5	83.3	407	10	Q8RXQ4	Q8rxq4 arabidopsis
8	4	66.7	52	16	Q97T74	Q97t74 streptococc
9	4	66.7	62	12	Q98153	Q98153 kaposi's sa
10	4	66.7	64	16	Q8R6N7	Q8r6n7 thermoanaer
11	4	66.7	76	10	Q94CT2	Q94ct2 oryza sativ
12	4	66.7	83	11	Q99LX1	Q99lx1 mus musculu
13	4	66.7	91	12	Q96812	Q96812 grapevine v
14	4	66.7	94	17	Q8U0Y1	Q8u0y1 pyrococcus
15	4	66.7	98	16	Q8U8F9	Q8u8f9 agrobacteri
16	4	66.7	102	16	Q8X2K1	Q8x2k1 escherichia
17	4	66.7	103	5	Q9XXT9	Q9xxt9 caenorhabdi
18	4	66.7	104	2	Q91BF3	Q91bf3 rhodospiril
19	4	66.7	106	2	Q9XCD1	Q9xcd1 thermomonos
20	4	66.7	106	17	Q8U1G4	Q8u1g4 pyrococcus
21	4	66.7	108	11	Q9D688	Q9d688 mus musculu
22	4	66.7	110	2	O86062	O86062 clostridiu
23	4	66.7	118	11	Q9CYT2	Q9cyt2 mus musculu
24	4	66.7	118	16	Q8Y1N5	Q8y1n5 ralstonia s
25	4	66.7	122	10	O24523	O24523 oryza sativ
26	4	66.7	127	4	Q9NXN1	Q9nxn1 homo sapien
27	4	66.7	129	10	Q8S6K5	Q8s6k5 oryza sativ
28	4	66.7	131	10	Q9M6T3	Q9m6t3 abies alba
29	4	66.7	131	10	Q9M6T1	Q9m6t1 abies alba
30	4	66.7	132	10	Q9CAE8	Q9cae8 arabidopsis
31	4	66.7	133	16	Q8UIF8	Q8uif8 agrobacteri
32	4	66.7	135	16	Q9PKQ4	Q9pkq4 chlamydia m
33	4	66.7	135	16	Q8Y151	Q8y151 ralstonia s
34	4	66.7	138	2	Q46009	Q46009 clostridium
35	4	66.7	143	11	Q8R1B5	Q8r1b5 mus musculu
36	4	66.7	145	2	O86437	O86437 pseudomonas
37	4	66.7	150	5	Q25036	Q25036 haliotis ry
38	4	66.7	150	5	Q25078	Q25078 haliotis la
39	4	66.7	150	5	Q25141	Q25141 haliotis ru
40	4	66.7	150	5	Q25048	Q25048 haliotis co
41	4	66.7	151	4	Q8WW34	Q8ww34 homo sapien
42	4	66.7	151	11	Q9DA47	Q9da47 mus musculu
43	4	66.7	154	5	Q25156	Q25156 haliotis ri
44	4	66.7	155	16	Q92NC6	Q92nc6 rhizobium m
45	4	66.7	161	4	Q9Y559	Q9y559 homo sapien

ALIGNMENTS

RESULT 1
O52770
ID O52770 PRELIMINARY; PRT; 301 AA.
AC O52770;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Activator protein.
GN MTLR.
OS Pseudomonas fluorescens.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=294;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DSM 50106;
RX MEDLINE=98332715; PubMed=9666063;
RA Brunker P., Hils M., Altenbuchner J., Mattes R.;
RT "The mannitol utilization genes of Pseudomonas fluorescens are
RT regulated by an activator: cloning, nucleotide sequence and expression
RT of the mtlR gene.";
RL Gene 215:19-27(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=DSM 50106;
RA Bruenker P., Altenbuchner J.;
RL Submitted (FEB-1998) to the EMBL/GenBank/DDBJ databases.
CC -!- SIMILARITY: BELONGS TO THE ARAC/XYLS FAMILY OF TRANSCRIPTIONAL
CC REGULATORS.
DR EMBL; AF047527; AAC34292.1; -.
DR InterPro; IPR000005; HTHAraC.
DR Pfam; PF00165; HTH_AraC; 2.
DR PRINTS; PR00032; HTHARAC.
DR SMART; SM00342; HTH_ARAC; 1
DR PROSITE; PS00041; HTH_ARAC_FAMILY_1; 1.
DR PROSITE; PS01124; HTH_ARAC_FAMILY_2; 1.
KW DNA-binding; Transcription regulation.
SQ SEQUENCE 301 AA; 34776 MW; 08D13FAF5524DDD6 CRC64;

```
  Query Match             93.3%;  Score 5;  DB 2;  Length 301;
  Best Local Similarity   100.0%;  Pred. No. 41;
  Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 VRWHF 6
           |||||
Db      35 VRWHF 39


RESULT 2
Q91.E1
ID   Q9I1E1      PRELIMINARY;      PRT;   301 AA.
AC   Q9I1E1;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Transcriptional regulator MtlR.
GN   MTLR OR PA2337.
OS   Pseudomonas aeruginosa.
OC   Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC   Pseudomonas.
OX   NCBI_TaxID=287;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 15692 / PAO1;
RX   MEDLINE=20437337; PubMed=10984043;
RA   Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warrener P.,
RA   Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA   Garber R.L., Goltry L., Tolentino E., Westbrock-Wadman S., Yuan Y.,
RA   Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA   Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA   Reizer J., Saier M.H., Hancock R.E.W., Lory S., Olson M.V.,
RT   "Complete genome sequence of Pseudomonas aeruginosa PAO1, an
RT   opportunistic pathogen.";
RL   Nature 406:959-964(2000).
CC   -!- SIMILARITY: BELONGS TO THE ARAC/XYLS FAMILY OF TRANSCRIPTIONAL
CC       REGULATORS.
DR   EMBL; AE004660; AAG05725.1; -.
DR   InterPro; IPR000005; HTHARAC.
DR   Pfam; PF00165; HTH_AraC; 2.
DR   PRINTS; PR00032; HTHARAC.
DR   SMART; SM00342; HTH_ARAC; 1.
DR   PROSITE; PS00041; HTH_ARAC_FAMILY_1; 1.
DR   PROSITE; PS01124; HTH_ARAC_FAMILY_2; 1.
KW   DNA-binding; Transcription regulation; Complete proteome.
SQ   SEQUENCE   301 AA;  34562 MW;  D8D0DB1EF7EEB1D CRC64;

  Query Match             93.3%;  Score 5;  DB 10;  Length 301;
  Best Local Similarity   100.0%;  Pred. No. 41;
  Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 VRWHF 6
           |||||
Db      35 VRWHF 39


RESULT 3
Q8WNV6
ID   Q8WNV6      PRELIMINARY;      PRT;   308 AA.
AC   Q8WNV6;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Putative pheromone receptor gV1R1.
GN   GV1R1
OS   Capra hircus (Goat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Caprinae; Capra.
OX   NCBI_TaxID=9925;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SHIBA;
RA   Wakabayashi Y., Mori Y., Ishihara M., Yaraki K., Hagino Yamagishi K.;
RT   "Gene of goat putative pheromone receptor is expressed in two distinct
RT   olfactory organs.";
RL   Submitted (JUN-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB064662; BAB83523.1; -.
DR   InterPro; IPR000276; GPCR_Rhodpsn.
DR   InterPro; IPR005136; V1R.
DR   Pfam; PF03402; V1R; 1.
DR   PROSITE; PS50262; G_PROTEIN_RECEP_F1_2; 1.
KW   Receptor.
SQ   SEQUENCE   308 AA;  35170 MW;  C8EA6B5F41420192 CRC64;

  Query Match             93.3%;  Score 5;  DB 6;  Length 308;
  Best Local Similarity   100.0%;  Pred. No. 41;
  Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 WVRWH 5
           |||||
Db     304 WVRWH 308


RESULT 4
Q9FC71
ID   Q9FC71      PRELIMINARY;      PRT;   366 AA.
AC   Q9FC71;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Hypothetical protein SCO7115.
GN   SCO7115 OR SC4B10.16C.
OS   Streptomyces coelicolor.
OC   Bacteria; Firmicutes, Actinobacteridae; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=1902;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2);
RA   Saunders D.C., Harris D.;
RL   Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2);
RA   Cerdeno A.M., Parkhill J., Barrell B.G., Rajandream M.A.;
RL   Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2);
RX   MEDLINE=97000301; PubMed=8843436;
RA   Redenbach M., Kieser H.M., Denapaite D., Eichner A., Cullum J.,
RA   Kinashi H., Hopwood D.A.;
RT   "A set of ordered cosmids and a detailed genetic and physical map for
RT   the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL   Mol. Microbiol. 21:77-96(1996).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2) / M145;
RA   Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
RA   Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
RA   Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA   Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA   Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neil S.,
RA   Rabbinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,
RA   Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA   Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
RA   Hopwood D.A.;
RT   "Complete genome sequence of the model actinomycete Streptomyces
RT   coelicolor A3(2).";
RL   Nature 417:141-147(2002).
DR   EMBL; AL391515; CAC04221.1; -.
DR   InterPro; IPR000873; AMP-bind.
DR   PROSITE; PS00455; AMP_BINDING; UNKNOWN_1.
KW   Hypothetical protein.
SQ   SEQUENCE   366 AA;  39766 MW;  2D9318B9B81F90B4 CRC64;
```

```
  Query Match             83.3%;  Score 5;  DB 15;  Length 366;
  Best Local Similarity   100.0%;  Pred. No. 47;
  Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 VRWHF 6
           |||||
Db      27 VRWHF 31


RESULT 5
Q9M9G2
ID   Q9M9G2      PRELIMINARY;      PRT;   370 AA.
AC   Q9M9G2;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   F14O23.22 protein.
GN   F14O23.22.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. COLUMBIA;
RA   Liu S., Yu G., Lee J., Sakano H., Jhaveri A., Lenz C., Toriumi M.,
RA   Chin C., Chiou J., Choi E., Gonzalez A., Howng P., Koo T., Li J.,
RA   Liu A., Pham P., Vaysberg M., Altafi H., Buehler E., Chao Q., Conn L.,
RA   Conway A.B., Hansen N., Johnson Hopson C., Khan S., Kim C., Lam P.,
RA   Nguyen M., Palm C., Shinn P., Tambunga G., Davis R., Ecker J.,
RA   Federspiel N., Theologis A.;
RT   "The sequence of BAC F14O23 from Arabidopsis thaliana chromosome 1.";
RL   Submitted (NOV-1999) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. COLUMBIA;
RA   Theologis A.;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: CONTAINS 7 WD REPEATS (TSP ASP DOMAINS).
DR   EMBL; AC012654; AAF43237.1; -.
DR   InterPro; IPR001680; WD40.
DR   Pfam; PF00400; WD40; 7.
DR   PRINTS; PR00320; GPROTEINBRPT.
DR   SMART; SM00320; WD40; 7.
DR   PROSITE; PS00678; WD_REPEATS_1; UNKNOWN_1.
DR   PROSITE; PS50082; WD_REPEATS_2; 7.
DR   PROSITE; PS50294; WD_REPEATS_REGION; 1.
KW   Repeat; WD repeat.
SQ   SEQUENCE   370 AA;  39774 MW;  CE58B01EB50C3B13 CRC64;

  Query Match             83.3%;  Score 5;  DB 10;  Length 370;
  Best Local Similarity   100.0%;  Pred. No. 48;
  Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 WVRWH 5
           |||||
Db     151 WVRWH 155


RESULT 6
O43898
ID   O43898      PRELIMINARY;      PRT;   376 AA.
AC   O43898;
DT   01-JUN-1998 (TrEMBLrel. 06, Created)
DT   01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   High-affinity lysophosphatidic acid receptor homolog.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BRAIN;
RA   An S.;
RL   Submitted (MAR-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U92642; AAB93884.1; -.
DR   InterPro; IPR000276; GPCR_Rhodpsn.
DR   Pfam; PF00001; 7tm_1; 1.
DR   PRINTS; PR00237; GPCRRHODOPSN.
DR   PROSITE; PS50262; G_PROTEIN_RECEP_F1_2; 1.
KW   Receptor.
SQ   SEQUENCE   376 AA;  41456 MW;  44C6E3B98CB830DF CRC64;

  Query Match             83.3%;  Score 5;  DB 4;  Length 376;
  Best Local Similarity   100.0%;  Pred. No. 48;
  Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 VRWHF 6
           |||||
Db      98 VRWHF 102


RESULT 7
Q8RXQ4
ID   Q8RXQ4      PRELIMINARY;      PRT;   407 AA.
AC   Q8RXQ4;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Hypothetical 43.8 kDa protein.
GN   AT1G71840.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Yamada K., Liu S.X., Sakano H., Pham P.K., Banh J., Chung M.K.,
RA   Goldsmith A.D., Lee J.M., Quach H.L., Toriumi M., Yu G., Bowser L.,
RA   Carninci P., Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T.,
RA   Kamiya A., Karlin-Neumann G., Kawai J., Kim C., Lam B., Lin J.,
RA   Miranda M., Narusaka M., Nguyen M., Palm C.J., Sakurai T., Satou M.,
RA   Seki M., Shinn P., Southwick A., Shinozaki K., Davis R.W., Ecker J.R.,
RA   Theologis A.;
RT   "Arabidopsis Full Length cDNA Clones.";
RL   Submitted (FEB-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY080722; AAL86002.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   407 AA;  43791 MW;  712D7CADD1790833 CRC64;

  Query Match             83.3%;  Score 5;  DB 10;  Length 407;
  Best Local Similarity   100.0%;  Pred. No. 51;
  Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 WVRWH 5
           |||||
Db     160 WVRWH 164


RESULT 8
Q97T74
ID   Q97T74      PRELIMINARY;      PRT;    52 AA.
AC   Q97T74;
DT   01-OCT-2001 (TrEMBLrel. 18, Created)
DT   01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   Hypothetical protein SP0077.
GN   SP0077.
OS   Streptococcus pneumoniae.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
OC   Streptococcaceae; Streptococcus.
OX   NCBI_TaxID=1313;
```

```
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=TIGR4;
RX   MEDLINE 21184910, PubMed 11544910,
RA   Tettelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,
RA   Peterson S., Heidelberg J., DeBoy R.T., Haft D.H., Dodson R.J.,
RA   Durkin A.S., Gwinn M., Kolonay J.F., Nelson W.C., Peterson J.D.,
RA   Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,
RA   Holtzapple E., Khouri H., Wolf A.M., Utterback T.R., Hansen C.L.,
RA   McDonald L.A., Feldblyum T.V., Angiuoli S., Dickinson T., Hickey E.K.,
RA   Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,
RA   Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.,
RT   "Complete genome sequence of a virulent isolate of Streptococcus
RT   pneumoniae.";
RL   Science 293:498-506(2001).
DR   EMBL; AE007323; AAK74268.1; -.
DR   TIGR; SP0077; -.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   82 AA;  9351 MW;  199464FF3749F34C CRC64;

  Query Match             66.7%;  Score 4;  DB 16;  Length 82;
  Best Local Similarity   100.0%;  Pred. No. 1.7e+02;
  Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 RWHF 6
           ||||
Db      19 RWHF 22


RESULT 9
Q98153
ID   Q98153      PRELIMINARY;      PRT;    62 AA.
AC   Q98153;
DT   01-FEB-1997 (TrEMBLrel. 02, Created)
DT   01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT   01-NOV 1999 (TrEMBLrel. 08, Last annotation update)
DE   Unidentified orf.
OS   Kaposi's sarcoma-associated herpesvirus (KSHV) (Human herpesvirus 8).
OC   Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC   Gammaherpesvirinae; Rhadinovirus.
OX   NCBI TaxID 37296;
RN   [1]
RP   SEQUENCE FROM N.A
RC   STRAIN=TUMOR;
RX   MEDLINE 98270695, PubMed 9600971;
RA   Zhong W., Wang H., Herndier B., Ganem D.;
RT   "Restricted expression of Kaposi sarcoma-associated herpesvirus (human
RT   herpesvirus 8) genes in Kaposi sarcoma.";
RL   Proc. Natl. Acad. Sci. U.S.A. 93:6641-6646(1996).
DR   EMBL; U66521; AAC55373.1; -.
SQ   SEQUENCE   62 AA;  7047 MW;  D7CAB11C1AA2BB2D CRC64;

  Query Match             66.7%;  Score 4;  DB 12;  Length 62;
  Best Local Similarity   100.0%;  Pred. No. 1e+02;
  Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 VRWH 5
           ||||
Db      31 VRWH 34


RESULT 10
Q8R6N7
ID   Q8R6N7      PRELIMINARY;      PRT;    61 AA.
AC   Q8R6N7;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-JUN 2002 (TrEMBLrel. 21, Last annotation update)
DE   Hypothetical protein TTE2763.
GN   TTE2763.
OS   Thermoanaerobacter tengcongensis.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
OC   Thermoanaerobacteriales; Thermoanaerobacteriaceae; Thermoanaerobacter.
OX   NCBI_TaxID=119072;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN MB4T / JCM11007;
RX   MEDLINE 21992916, PubMed 11997336;
RA   Bao Q., Tian Y., Li W., Xu Z., Xuan Z., Hu S., Dong W., Yang J.,
RA   Chen Y., Xue Y., Xu Y., Lai X., Huang L., Ding X., Ma Y., Ling L.,
RA   Tan H., Chen R., Wang J., Yu J., Yang H.;
RT   "A complete sequence of T. tengcongensis genome.";
RL   Genome Res. 12:689-700(2002).
DR   EMBL; AE013214; AAM25868.1; -.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   64 AA;  7330 MW;  AC6B35D798041CF4 CRC64;

  Query Match             66.7%;  Score 4;  DB 16;  Length 64;
  Best Local Similarity   100.0%;  Pred. No. 2e+02;
  Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 RWHF 6
           ||||
Db      56 RWHF 59


RESULT 11
Q94CT2
ID   Q94CT2      PRELIMINARY;      PRT;    76 AA.
AC   Q94CT2;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   P0459B04.14 protein.
GN   P0459B04.14.
OS   Oryza sativa (Rice).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC   Ehrhartoideae; Oryzeae; Oryza.
OX   NCBI_TaxID=4530;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN CV. NIPPONBARE;
RA   Sasaki T., Matsumoto T., Yamamoto K.;
RT   "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
RT   clone:P0459B04.";
RL   Submitted (MAY-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AP003607; BAB62830.1; -.
SQ   SEQUENCE   76 AA;  8274 MW;  15BAF5C63CD38EDA CRC64;

  Query Match             66.7%;  Score 4;  DB 16;  Length 76;
  Best Local Similarity   100.0%;  Pred. No. 2.3e+02;
  Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 RWHF 6
           ||||
Db      15 RWHF 18


RESULT 12
Q99LX1
ID   Q99LX1      PRELIMINARY;      PRT;    83 AA.
AC   Q99LX1;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   H,p-tubulin 3.5 kDa protein (Fragment).
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Euarchontoglires; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Strausberg R.;
RL   Submitted (JAN 2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC002185; AAH02185.1; -.
```

```
KW   Hypothetical protein.
FT   NON_TER       1      1
SQ   SEQUENCE   83 AA;  9486 MW;  5C9079B81C8941ED CRC64;

  Query Match             66.7%;  Score 4;  DB 11;  Length 83;
  Best Local Similarity   100.0%;  Pred. No. 2.5e+02;
  Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 WVRW 4
           ||||
Db      54 WVRW 57


RESULT 13
Q96812
ID   Q96812      PRELIMINARY;      PRT;    91 AA.
AC   Q96812;
DT   01-FEB-1997 (TrEMBLrel. 02, Created)
DT   01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   RNA binding protein.
OS   Grapevine virus D.
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Vitivirus
OX   NCBI_TaxID=51617;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Abou Ghanem N., Saldarelli P., Minafra A., Buzkan N., Castellano M.A.,
RA   Martelli G.P.;
RT   "Properties of grapevine virus D, a novel putative trichovirus.";
RL   J. Plant Pathol. 79:15-25(1997).
DR   EMBL; Y07764; CAA69071.1; -.
SQ   SEQUENCE   91 AA;  10459 MW;  7B3BA2F15D0249B8 CRC64;

  Query Match             66.7%;  Score 4;  DB 12;  Length 91;
  Best Local Similarity   100.0%;  Pred. No. 2.7e+02;
  Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 WVRW 4
           ||||
Db      70 WVRW 73


RESULT 14
Q8U0Y1
ID   Q8U0Y1      PRELIMINARY;      PRT;    94 AA.
AC   Q8U0Y1;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Hypothetical protein PF1450.
GN   PF1450.
OS   Pyrococcus furiosus.
OC   Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
OC   Pyrococcus.
OX   NCBI_TaxID=2261;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=VC1 / DSM 3638 / ATCC 43587 / JCM 8422;
RA   Weiss R.B., Dunn D.M., Robb F.T., Brown J.R.;
RT   "The complete sequence of the Pyrococcus furiosus genome.";
RL   Submitted (FEB-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AE010247; AAL81574.1; -.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   94 AA;  10554 MW;  CE8AD48A76AF7CE8 CRC64;

  Query Match             66.7%;  Score 4;  DB 17;  Length 94;
  Best Local Similarity   100.0%;  Pred. No. 2.8e+02;
  Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 WVRW 4
           ||||
Db      91 WVRW 94


RESULT 15
Q8U8F9
ID   Q8U8F9      PRELIMINARY;      PRT;    98 AA.
AC   Q8U8F9;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Hypothetical protein Atu4133.
GN   ATU4133.
OS   Agrobacterium tumefaciens (strain C58 / ATCC 33970).
OC   Bacteria, Proteobacteria; alpha subdivision, Rhizobiaceae group;
OC   Rhizobiaceae; Rhizobium.
OX   NCBI_TaxID=176299;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21608550; PubMed=11743193;
RA   Wood D.W., Setubal J.C., Kaul R., Monks D.E., Kitajima J.P.,
RA   Okura V.K., Zhou Y., Chen L., Wood G.E., Almeida N.F. Jr.,
RA   Chen Y., Paulsen I.T., Eisen J.A., Karp P.D., Bovee D. Sr.,
RA   Chapman P., Clendenning J., Deatherage G., Gillet W., Grant C.,
RA   Kutyavin T., Levy R., Li M.-J., McClelland E., Palmieri A.,
RA   Raymond C., Rouse G., Saenphimmachak C., Wu Z., Romero P.,
RA   Zhang G., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Perry M.,
RA   Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.Y., Dolan M.,
RA   Chumley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,
RA   Nester E.W.;
RT   "The genome of the natural genetic engineer Agrobacterium tumefaciens
RT   C58.";
RL   Science 294:2317-2323(2001).
DR   EMBL; AE009344; AAL44933.1; -.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   98 AA;  10925 MW;  8362B6F5942B3018 CRC64;

  Query Match             66.7%;  Score 4;  DB 16;  Length 98;
  Best Local Similarity   100.0%;  Pred. No. 2.9e+02;
  Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 VRWH 5
           ||||
Db       3 VRWH 6


Search completed: February 20, 2003, 11:28:56
Job time : 47 secs
```

GenCore version 5.1.3

OM protein - protein search, using sw model

Run on: February 20, 2003, 10:35:40 ; Search time 29.3333 Seconds
(without alignments)
27.256 Million cell updates/sec

Title: US-09-912-414-2
Perfect score: 45
Sequence: 1 WVPWHF 6

Scoring table: BLOSUM62DX
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	45	100.0	6	21	AAB01492	Peptide which bind
2	41	91.1	6	21	AAB01505	Peptide which bind
3	41	91.1	256	21	AAG09463	Arabidopsis thalia
4	41	91.1	256	21	AAG44075	Arabidopsis thalia
5	41	91.1	256	21	AAG48293	Arabidopsis thalia
6	41	91.1	325	21	AAG09462	Arabidopsis thalia
7	41	91.1	325	21	AAG44074	Arabidopsis thalia
8	41	91.1	325	21	AAG48292	Arabidopsis thalia
9	40	88.9	306	21	AAG05810	Arabidopsis thalia
10	40	88.9	306	21	AAG47194	Arabidopsis thalia
11	40	88.9	338	21	AAG05809	Arabidopsis thalia
12	40	88.9	338	21	AAG47193	Arabidopsis thalia
13	40	88.9	418	15	AAR60501	Linoleic acid desa
14	40	88.9	448	21	AAG05808	Arabidopsis thalia
15	40	88.9	448	21	AAG47192	Arabidopsis thalia
16	40	88.9	448	23	ABB92179	Herbicidally activ
17	39	86.7	6	21	AAB01506	Peptide which bind
18	39	86.7	6	21	AAB01507	Peptide which bind
19	38	84.4	477	21	AAY51600	Human truncated tr
20	38	84.4	822	16	AAR81630	Human trkB recepto
21	38	84.4	822	21	AAY51599	Human trkB recepto
22	38	84.4	822	23	AAM50851	Human receptor tyr
23	36	80.0	79	21	AAR40077	Human CRFX CRF41 p
24	36	80.0	79	23	ABP11225	Human CRFX protein
25	36	80.0	83	20	AAY25731	Human secreted pro
26	36	80.0	91	23	ABP07463	Novel human diagno
27	36	80.0	424	22	AAE13431	Soybean microsomal
28	36	80.0	1169	22	ABB71663	Drosophila melanog
29	35	77.8	6	21	AAB01508	Peptide which bind
30	35	77.8	53	23	ABG64802	Human albumin fusi
31	35	77.8	53	23	AAE21233	Human gene 18 enco
32	35	77.8	54	22	AAO09345	Human polypeptide
33	35	77.8	148	22	ABB11414	Human breast cance
34	35	77.8	173	22	AAG80884	Lipid modification
35	35	77.8	173	22	AAG83334	P patens lipid met
36	35	77.8	663	22	AAU51530	Propionibacterium
37	34	75.6	6	21	AAR01504	Peptide which bind
38	34	75.6	8	15	AAR60429	Antiproliferative
39	34	75.6	8	15	AAR60444	Antiproliferative
40	34	75.6	24	21	AAB53827	Human colon cancer
41	34	75.6	42	22	AAU14682	Novel bone marrow
42	34	75.6	49	22	AAM91182	Human immune/haema
43	34	75.6	80	22	AAM95613	Human immune/haema
44	34	75.6	90	22	AAU14777	Novel bone marrow
45	34	75.6	178	22	AAU44711	Propionibacterium

ALIGNMENTS

RESULT 1
AAB01492
ID AAB01492 standard; peptide; 6 AA.
XX
AC AAB01492;
XX
DT 08-NOV-2000 (first entry)
XX
DE Peptide which binds to transcription factor E2F 1 DNA binding domain.
XX
KW DNA binding, transcription factor, E2F, E2F 1, cell cycle; DP 1;
KW activation, transcription; apoptosis; proliferative disorder;
KW psoriasis; restenosis.
XX
OS Synthetic.
XX
PN WO200044771-A1.
XX
PD 03-AUG-2000.
XX
PF 26-JAN-2000; 2000WO-GB00227.
XX
PR 26-JAN-1999; 99GB-0001710.
XX
PA (PROL-) PROLIFIX LTD.
XX
PI Mueller R, Kontermann RE, Montigiani S;
XX
DR WPI; 2000-532806/48.
XX
PT Peptides binding to the DNA binding domain of transcription factor E2F
PT and inhibiting cell cycle progression, useful for the treatment of
PT cancer

XX
PS Claim 6; Page 2; 42pp; English.
XX
CC Peptides which bind to the DNA binding domain of transcription
CC factor E2F and inhibit cell cycle progression may be useful as
CC research agents to investigate the interaction between E2F and DP-1,
CC or the activation of transcription by E2F 1/DP 1 heterodimers. They
CC may also be used for inducing apoptosis and/or cell cycle arrest in
CC a cell, particularly for treatment of cancer or other proliferative
CC disorders such as psoriasis and restenosis.
XX
SQ Sequence 6 AA;

Query Match 100.0%; Score 45; DB 21; Length 6;
Best Local Similarity; 100.0%; Pred. No. 7.8e-3;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 WVRWHF 6
||||||
Db 1 WVRWHF 6

RESULT 2
AAB01505
ID AAB01505 standard; peptide; 6 AA.
XX
AC AAB01505;
XX
DT 08-NOV-2000 (first entry)
XX
DE Peptide which binds to transcription factor E2F 1 DNA binding domain.
XX
KW DNA binding; transcription factor 1; E2F; E2F 1; cell cycle; cell
KW activation; transcription; apoptosis; proliferative disorder;
KW psoriasis; restenosis.
XX
OS Synthetic
XX
PN WO200044771-A1.
XX
PD 03-AUG-2000.
XX
PF 26-JAN-2000; 2000WO-GB00222.
XX
PR 26-JAN-1999; 99GB-0001710.
XX
PA (PROL-) PROLIFIX LTD.
XX
PI Mueller R, Kuhlemann RE, Montigiani S,
XX
DR WPI; 2000-532806/48.
XX
PT Peptides binding to the DNA binding domain of transcription factor E2F
PT and inhibiting cell cycle progression, useful for the treatment of
PT cancer
XX
PS Example; Page 26; 42pp; English.
XX
CC Peptides which bind to the DNA binding domain of transcription
CC factor E2F and inhibit cell cycle progression may be useful as
CC research agents to investigate the interaction between E2F and DP-1,
CC or the activation of transcription by E2F 1/DP 1 heterodimers. They
CC may also be used for inducing apoptosis and/or cell cycle arrest in
CC a cell, particularly for treatment of cancer or other proliferative
CC disorders such as psoriasis and restenosis.
XX
SQ Sequence 6 AA;

Query Match 91.1%; Score 41; DB 21; Length 6;
Best Local Similarity; 83.3%; Pred. No. 7.8e-03;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 WVRWHF 6
|||| |
Db 1 WARWHF 6

RESULT 3
AAG09463
ID AAG09463 standard; protein; 256 AA.
XX
AC AAG09463;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO 7409.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 05-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 31-MAR-1999; 99US-0127162.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132497.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134769.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136332.
PR 28-MAY-1999; 99US-0136382.
PR 01-JUN-1999; 99US-0137022.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137582.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 15-JUN-1999; 99US-0139402.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139464.

PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144362.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158032.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 91.1%; Score 41; DB 21; Length 256;
Best Local Similarity 83.3%; Pred. No. 35,
Matches 5; Conservative 0, Mismatches 1; Indels 0; Gaps 0;

```
Qy       1 WVRWHF 6
             | ||||
Db      72 WARWHF 77
```

RESULT 4
AAG44075
ID AAG44075 standard; Protein; 256 AA.

XX
AC AAG44075;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO 5164.
XX
KW Protein identification, signal transduction pathway, metabolic pathway,
KW hybridisation assay, genetic mapping, gene expression control, promoter,
KW termination sequence.
XX
OS Arabidopsis thaliana
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123160.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788
PR 25-MAR-1999; 99US-0126264
PR 29-MAR-1999; 99US-0126788
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234
PR 08-APR-1999; 99US-0128714
PR 16-APR-1999; 99US-0129845
PR 19-APR-1999; 99US-0130077
PR 21-APR-1999; 99US-0130543
PR 23-APR-1999; 99US-0130519
PR 23-APR-1999; 99US-0130881
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484
PR 05-MAY-1999; 99US-0132485
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132864.
PR 11-MAY-1999; 99US-0134256
PR 14-MAY-1999; 99US-0134218
PR 14-MAY-1999; 99US-0134219
PR 14-MAY-1999; 99US-0134221
PR 14-MAY-1999; 99US-0134370
PR 18-MAY-1999; 99US-0134769.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124
PR 21-MAY-1999; 99US-0135354
PR 24-MAY-1999; 99US-0135629
PR 25-MAY-1999; 99US-0136021
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN 1999; 99US-0137602.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847
PR 14-JUN-1999; 99US-0139119
PR 16-JUN-1999; 99US-0139452
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139442
PR 18-JUN 1999; 99US-0139454
PR 18-JUN-1999; 99US-0139455
PR 18-JUN-1999; 99US-0139456
PR 18-JUN-1999; 99US-0139457
PR 18-JUN 1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN 1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461

PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139879.
PR 23-JUN 1999; 99US-0140353.
PR 23-JUN 1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140485.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12 JUL 1999; 99US 0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16 JUL 1999; 99US-0144096.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19 JUL 1999; 99US 0144334.
PR 19 JUL 1999; 99US 0144335.
PR 20-JUL-1999; 99US-0144552.
PR 20-JUL-1999; 99US-0144682.
PR 20-JUL-1999; 99US 0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL 1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22 JUL 1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22 JUL 1999; 99US 0145192.
PR 23 JUL 1999; 99US 0145145.
PR 23-JUL-1999; 99US-0145210.
PR 23-JUL-1999; 99US-0145224.
PR 26 JUL 1999; 99US 0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27 JUL 1999; 99US 0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02 AUG 1999; 99US 0146296.
PR 02 AUG 1999; 99US-0146298.
PR 02-AUG-1999; 99US-0146389.
PR 03 AUG 1999; 99US 0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG 1999; 99US-0147302.
PR 05 AUG 1999; 99US 0147192.
PR 05 AUG 1999; 99US 0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG 1999; 99US 0147935.
PR 10 AUG 1999; 99US 0148171.
PR 11 AUG 1999; 99US 0148319.
PR 12 AUG 1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148694.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20 AUG 1999; 99US 0149702.
PR 20 AUG 1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23 AUG 1999; 99US 0149930.

PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156459.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 91.1%; Score 41; DB 21; Length 256;
Best Local Similarity 83.3%; Pred. No. 35;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy      1 WVRWHF 6
          | ||||
Db     72 WARWHF 77
```

RESULT 5
AAG48293
ID AAG48293 standard; Protein; 256 AA.
XX
AC AAG48293;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 60969
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PP 25-FEB-1999; 99US-0121825.
PP 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PP 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PP 16-APR-1999; 99US-0129845.
PP 19-APR-1999; 99US-0130077.
PP 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PP 30-APR-1999; 99US-0132048.
PP 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PP 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PP 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PP 18-MAY-1999; 99US-0134768.
PP 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PP 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PP 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PP 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18 JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PP 22 JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.

PR 23-JUN-1999; 99US-0140254.
PR 24-JUN-1999; 99US-0140635.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0141854.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142296.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142922.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144225.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144634.
PR 20-JUL-1999; 99US-0144694.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145084.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145219.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145478.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145915.
PR 27-JUL-1999; 99US-0145917.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146385.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0146939.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147432.
PR 05-AUG-1999; 99US-0147460.
PR 06-AUG-1999; 99US-0147482.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150500.
PR 25-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152263.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158252.
PR 12-OCT-1999; 99US-0158360.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160984.
PR 25-OCT-1999; 99US-0161104.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161514.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 31.1%, Score 41, DB 21, Length 256;
Best Local Similarity 93.3%, Pred. No. 35;
Matches 7, Conservative 0, Mismatches 1, Indels 0, Gaps 0;

Qy 1 WVRWHF 6
|||||| |
Db 72 WARWHF 77

RESULT 6
AAG09462
ID AAG09462 standard; Protein; 325 AA.
XX
AC AAG09462;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 7408.
XX
KW Protein identification, signal transduction pathway, metabolic pathway;
KW hybridisation assay, genetic mapping, gene expression control, promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.

XX
PD 06-SEP-2000
XX
PF 25-FEB-2000; 2000EP-0301479
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125798.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US 0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484
PR 05-MAY-1999; 99US-0132485
PR 06-MAY-1999; 99US-0132486
PR 06-MAY-1999; 99US-0132487
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256
PR 14-MAY-1999; 99US-0134218
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941
PR 20-MAY-1999; 99US-0135124
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459
PR 18-JUN-1999; 99US-0139460
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695
PR 28-JUN-1999; 99US 0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086
PR 19-JUL-1999; 99US-0144325
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US 0145132.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US 0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US 0149722.
PR 20 AUG 1999; 99US-0149723.
PR 20-AUG-1999; 99US 0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930
PR 07 SEP 1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.

PR 22-SEP-1999; 99US-0155154.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155864.
PR 28-SEP-1999; 99US-0156456.
PR 29-SEP-1999; 99US-0156698.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157985.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158222.
PR 12-OCT-1999; 99US-0158365.
PR 13-OCT-1999; 99US-0159264.
PR 13-OCT-1999; 99US-0159284.
PR 13-OCT-1999; 99US-0159285.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159367.
PR 14-OCT-1999; 99US-0159455.
PR 14-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160784.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 26-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 91.1%; Score 41; DB 21; Length 325;
Best Local Similarity 83.3%; Pred. No. 45;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 WVPWHP 6
| .|||
Db 141 WARWHF 146

RESULT 7
AAG44074
ID AAG44074 standard; Protein; 331 AA.
XX
AC AAG44074;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO 55102.
XX
KW Protein identification, signal transduction pathway, metabolic pathway,
KW hybridisation assay, genetic mapping, gene expression, DNA, in silico,
KW termination sequence.
XX
OS Arabidopsis thaliana
XX
PN EP1033405-A2
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439
XX
PR 26-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.

PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0126788
PR 25-MAR-1999; 99US-0126864.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136792.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137548.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140923.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141942.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142590.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.

PR 15-JUL-1999; 99US-0144005
PR 16-JUL-1999; 99US-0144085
PR 16-JUL-1999; 99US-0144086
PR 19-JUL-1999; 99US-0144325
PR 19-JUL-1999; 99US-0144331
PR 19-JUL-1999; 99US-0144332
PR 19-JUL-1999; 99US-0144333
PR 19-JUL 1999; 99US-0144334
PR 19-JUL-1999; 99US-0144335
PR 20-JUL-1999; 99US-0144352
PR 20-JUL-1999; 99US-0144632
PR 20-JUL-1999; 99US-0144884
PR 21-JUL-1999; 99US-0144814
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL 1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089
PR 22-JUL-1999; 99US-0145192
PR 23-JUL-1999; 99US-0145145
PR 23-JUL-1999; 99US-0145218
PR 23-JUL-1999; 99US-0145224
PR 26-JUL-1999; 99US-0145276
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918
PR 27-JUL-1999; 99US-0145919
PR 28-JUL-1999; 99US-0146051
PR 02-AUG-1999; 99US-0146386
PR 02-AUG-1999; 99US-0146388
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038
PR 04-AUG-1999; 99US-0147204
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149369
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566
PR 26-AUG-1999; 99US-0150884
PR 27-AUG-1999; 99US-0151065
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039
PR 20-SEP-1999; 99US-0154779
PR 22-SEP-1999; 99US-0155139
PR 23-SEP-1999; 99US-0155486
PR 24-SEP-1999; 99US-0155659
PR 28-SEP-1999; 99US-0156458
PR 29-SEP-1999; 99US-0156596
PR 04-OCT-1999; 99US-0157117
PR 05-OCT-1999; 99US-0157253
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981
PR 22-OCT-1999; 99US-0160989
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 91.1%; Score 41; DB 21; Length 325;
Best Local Similarity 83.3%; Pred. No. 45;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 WVRWHF 6
| ||||
Db 141 WARWHF 146

RESULT 8
AAG48292
ID AAG48292 standard; Protein; 325 AA.
XX
AC AAG48292;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 60968.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405 A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.

PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130442.
PR 23-APR-1999; 99US-0130811.
PR 23-APR-1999; 99US-0130991.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132861.
PR 11-MAY-1999; 99US-0133256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134278.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135251.
PR 24-MAY-1999; 99US-0135600.
PR 25-MAY-1999; 99US-0135721.
PR 27-MAY-1999; 99US-0135892.
PR 28-MAY-1999; 99US-0136282.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139462.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139760.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140645.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142260.
PR 06-JUL-1999; 99US-0142713.
PR 08-JUL-1999; 99US-0142863.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0143277.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.

PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144432.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 03-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147147.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US 0148341.
PR 13-AUG-1999; 99US 0148553.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 15-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US 0154779.
PR 22-SEP-1999; 99US-0155139.
PR 22-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US 0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158252.
PR 12-OCT-1999; 99US-0159369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.

PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159537.
PR 14-OCT-1999; 99US-0159538.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160880.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 91.1%; Score 41; DB 21; Length 325;
Best Local Similarity 83.3%; Pred. No. 45;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 WVRWHF 6
| ||||
Db 141 WARWHF 146

RESULT 9
AAG05810
ID AAG05810 standard; Protein; 306 AA
XX
AC AAG05810;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 2254
XX
KW Protein identification; signal transduction pathway, metabolic pathway,
KW hybridisation assay, genetic mapping, gene expression control, promoter,
KW termination sequence
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144382.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.

PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145275.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146985.
PR 02-AUG-1999; 99US-0146988.
PR 02-AUG-1999; 99US-0146989.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147283.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 26-AUG-1999; 99US-0150866.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155159.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155654.
PR 29-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157353.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158024.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158569.
PR 13-OCT-1999; 99US-0158793.
PR 13-OCT-1999; 99US-0158794.
PR 13-OCT-1999; 99US-0158795.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159357.
PR 14-OCT-1999; 99US-0159359.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160983.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 98.9%; Score 46; DB 21; Length 306;
Best Local Similarity 83.3%; Pred. No. 60;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 WVRWHF 6
|| |:|
Db 115 WVNWHF 120

RESULT 10
AAG47194
ID AAG47194 standard; Protein; 306 AA.
XX
AC AAG47194;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO. 59455.
XX
KW protein identification, signal transduction pathway, metabolic pathway;
KW hybridization assay, protein mapping, gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 26-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 08-MAR-1999; 99US-0123548.
PR 25-MAR-1999; 99US-0125788.
PR 30-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 05-APR-1999; 99US-0128234.
PR 09-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891
PR 26-APR-1999; 99US-0131443.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.

PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137022.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139839.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145132.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.

PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147132.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.

PR 22-OCT-1999; 99US-0140989.
PR 22-OCT-1999; 99US-0141464.
PR 25-OCT-1999; 99US-0141405.
PR 25-OCT-1999; 99US-0141426.
PR 26-OCT-1999; 99US-0141364
PR 26-OCT-1999; 99US-0141360.
PR 26-OCT-1999; 99US-0141361.
PR 28-OCT-1999; 99US-0141320.
PR 28-OCT-1999; 99US-0141992.
PR 28-OCT-1999; 99US-0141994.
PR 29-OCT-1999; 99US-0142142

Query Match 99.9%; Score 4[illegible]; DB 21; Length 4[illegible];
Best Local Similarity 83.3%; Pred. No. 60;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 WVRWHF 6
|| |||
Db 116 WVNWHF 120

RESULT 11
AAG05809
ID AAG05809 standard; Protein; 338 AA.
XX
AC AAG05809;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO 2[illegible]
XX
KW Protein identification, signal transduction pathway, metabolic pathway,
KW hybridisation assay, genetic mapping, gene expression control, promoter,
KW termination sequence.
XX
OS Arabidopsis thaliana
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301434
XX
PR 25-FEB-1999; 99US-0121925.
PR 05-MAR-1999; 99US-0123180
PR 09-MAR-1999; 99US-0123548
PR 23-MAR-1999; 99US-0125788
PR 25-MAR-1999; 99US-0126264
PR 29-MAR-1999; 99US-0126795
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449
PR 23-APR-1999; 99US-0130610
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048
PR 30-APR-1999; 99US-0132407
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768
PR 19-MAY-1999; 99US-0134941
PR 20-MAY-1999; 99US-0135124.

PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135827.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137224.
PR 03-JUN-1999; 99US-0137538.
PR 04-JUN-1999; 99US-0137560.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140364.
PR 24-JUN-1999; 99US-0140605.
PR 26-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145376.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 29-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.

PR 01-AUG-1999; 99US-0147089.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155135.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156459.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159639.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160999.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.

PP 28-OCT-1999; 99US-0161920.
PP 28-OCT-1999; 99US-0161992.
PP 28-OCT-1999; 99US-0161993.
PP 29-OCT-1999; 99US-0162142.

Query Match 88.3%; Score 40; DB 21; Length 338;
Best Local Similarity 83.3%; Pred. No. 67;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 WVRWHF 6
|| |||
Db 147 WVNWHF 152

RESULT 12
AAG47193
ID AAG47193 standard; Protein; 338 AA
XX
AC AAG47193;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 59454.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137220.
PR 03-JUN-1999; 99US-0137528.

PR 04-JUN-1999; 99US-0137702.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138294.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139115.
PR 16-JUN-1999; 99US-0139452.
PR 17-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139464.
PR 18-JUN-1999; 99US-0139465.
PR 18-JUN-1999; 99US-0139466.
PR 18-JUN-1999; 99US-0139467.
PR 18-JUN-1999; 99US-0139468.
PR 18-JUN-1999; 99US-0139469.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139751.
PR 21-JUN-1999; 99US-0140917.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140363.
PR 23-JUN-1999; 99US-0140364.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140923.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141087.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142184.
PR 02-JUL-1999; 99US-0142065.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144305.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144804.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 23-JUL-1999; 99US-0145102.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 26-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145917.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 01-AUG-1999; 99US-0146765.
PR 02-AUG-1999; 99US-0146888.
PR 02-AUG-1999; 99US-0146983.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147206.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147190.
PR 05-AUG-1999; 99US-0147260.

PR 06-AUG-1999; 99US-0147203.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147835.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148694.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 19-AUG-1999; 99US-0149700.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151380.
PR 30-AUG-1999; 99US-0151702.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156469.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157353.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0159369.
PR 13-OCT-1999; 99US-0159283.
PR 13-OCT-1999; 99US-0159284.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161420.
PR 28-OCT-1999; 99US-0161900.
PR 29-OCT-1999; 99US-0161903.
PR 29-OCT-1999; 99US-0162142.

Query Match 99.3%; Score 43; DB 21; Length 338;
Best Local Similarity 99.3%; Pred. No. 67;

```
  Matches    5;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 WVRWHF 6
           || |||
Db     147 WVNWHF 152


RESULT 13
AAR60501
ID   AAR60501 standard; Protein; 418 AA.
XX
AC   AAR60501;
XX
DT   28-MAR-1995  (first entry)
XX
DE   Linoleic-acid-desaturase
XX
KW   linoleic-acid-desaturase; oleic-desaturase; transgenic plant;
KW   crop improvement; linolenic acid.
XX
OS   Not specified.
XX
PN   WO9418337-A.
XX
PD   18-AUG-1994.
XX
PF   04-FEB-1994;   94WO-US01321
XX
PR   05-FEB-1993;   93US-0014431.
PR   22-NOV-1993;   93US-0156561.
XX
PA   (MONS ) MONSANTO CO
PA   (UNMS ) UNIV MICHIGAN STATE
XX
PI   Arondel VJA,  Gibson SI,  Kishore GM,  Puff TG,  Somerville CR;
XX
DR   WPI; 1994-279758/34.
DR   N-PSDB; AAQ71242.
XX
PT   Genetically transformed plants with altered linolenic acid
PT   content - contg recombinant, double stranded DNA encoding
PT   linoleic acid desaturase, or the antisense of the coding
PT   sequence
XX
PS   Disclosure; Page 100-102; 144pp; English
XX
CC   The isolation of cDNA encoding linoleic-acid-desaturase and
CC   oleic-desaturase is described.  A sequence of the invention is
CC   given in AAQ71242, and its encoded protein sequence in AAR60501
XX
SQ   Sequence   418 AA;

  Query Match             88.9%;  Score 40;  DB 15;  Length 418;
  Best Local Similarity   83.3%;  Pred. No. 82;
  Matches    5;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 WVRWHF 6
           || |||
Db     257 WVNWHF 262


RESULT 14
AAG05808
ID   AAG05808 standard; Protein; 448 AA.
XX
AC   AAG05808;
XX
DT   17-OCT-2000  (first entry)
XX
DE   Arabidopsis thaliana protein fragment SEQ ID NO: 2352
XX
KW   Protein identification; signal transduction pathway; metabolic pathway,
KW   hybridisation assay, genetic mapping, gene expression control, promoter,
KW   termination sequence.
XX
OS   Arabidopsis thaliana
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.
XX
PR   25-FEB-1999;   99US-0121825.
PR   05-MAR-1999;   99US-0123180.
PR   09-MAR-1999;   99US-0123548.
PR   23-MAR-1999;   99US-0125788.
PR   25-MAR-1999;   99US-0126264.
PR   26-MAR-1999;   99US-0126785.
PR   01-APR-1999;   99US-0127462.
PR   06-APR-1999;   99US-0128234.
PR   08-APR-1999;   99US-0128714.
PR   16-APR-1999;   99US-0129845.
PR   19-APR-1999;   99US-0130077.
PR   21-APR-1999;   99US-0130449.
PR   23-APR-1999;   99US-0130510.
PR   23-APR-1999;   99US-0130891.
PR   28-APR-1999;   99US-0131449.
PR   30-APR-1999;   99US-0132048.
PR   30-APR-1999;   99US-0132407.
PR   04-MAY-1999;   99US-0132484.
PR   05-MAY-1999;   99US-0132485.
PR   06-MAY-1999;   99US-0132486.
PR   06-MAY-1999;   99US-0132487.
PR   07-MAY-1999;   99US-0132863.
PR   11-MAY-1999;   99US-0134256.
PR   14-MAY-1999;   99US-0134218.
PR   14-MAY-1999;   99US-0134219.
PR   14-MAY-1999;   99US-0134221.
PR   14-MAY-1999;   99US-0134370.
PR   18-MAY-1999;   99US-0134768.
PR   19-MAY-1999;   99US-0134941.
PR   20-MAY-1999;   99US-0135134.
PR   21-MAY-1999;   99US-0135353.
PR   24-MAY-1999;   99US-0135629.
PR   25-MAY-1999;   99US-0136021.
PR   27-MAY-1999;   99US-0136392
PR   28-MAY-1999;   99US-0136782
PR   01-JUN-1999;   99US-0137222
PR   03-JUN-1999;   99US-0137528.
PR   04-JUN-1999;   99US-0137502
PR   07-JUN-1999;   99US-0137724
PR   08-JUN-1999;   99US-0138094.
PR   10-JUN-1999;   99US-0138540.
PR   10-JUN-1999;   99US-0138847.
PR   14-JUN-1999;   99US-0139119.
PR   16-JUN-1999;   99US-0139452.
PR   16-JUN-1999;   99US-0139453.
PR   17-JUN-1999;   99US-0139492.
PR   18-JUN-1999;   99US-0139454.
PR   18-JUN-1999;   99US-0139455.
PR   18-JUN-1999;   99US-0139456.
PR   18-JUN-1999;   99US-0139457.
PR   18-JUN-1999;   99US-0139458.
PR   18-JUN-1999;   99US-0139459.
PR   18-JUN-1999;   99US-0139460.
PR   18-JUN-1999;   99US-0139461.
PR   18-JUN-1999;   99US-0139462.
PR   18-JUN-1999;   99US-0139463.
PR   18-JUN-1999;   99US-0139750.
PR   18-JUN-1999;   99US-0139763.
PR   21-JUN-1999;   99US-0139817.
PR   22-JUN-1999;   99US-0139899.
PR   23-JUN-1999;   99US-0140353.
PR   23-JUN-1999;   99US-0140354.
PR   24-JUN-1999;   99US-0140695.
```

PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141087.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142340.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142929.
PR 12-JUL-1999; 99US-0143077.
PR 12-JUL-1999; 99US-0143647.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144085.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144684.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145102.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 26-JUL-1999; 99US-0145241.
PR 26-JUL-1999; 99US-0145270.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145919.
PR 27-JUL-1999; 99US-0145919.
PR 29-JUL-1999; 99US-0146061.
PR 02-AUG-1999; 99US-0146384.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147294.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PP 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149269.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149725.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149932.
PR 23-AUG-1999; 99US-0149930.
PP 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150954.
PR 27-AUG-1999; 99US-0151065.
PP 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.

PR 10-SEP-1999; 99US-0153070.
PP 13-SEP-1999; 99US-0153758.
PP 15-SEP-1999; 99US-0154019.
PP 16-SEP-1999; 99US-0154039.
PP 20-SEP-1999; 99US-0154779.
PP 22-SEP-1999; 99US-0155139.
PP 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PP 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PP 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PP 08-OCT-1999; 99US-0158232.
PP 12-OCT-1999; 99US-0159079.
PP 12-OCT-1999; 99US-0159293.
PP 12-OCT-1999; 99US-0159294.
PP 13-OCT-1999; 99US-0159295.
PP 14-OCT-1999; 99US-0159329.
PP 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PP 14-OCT-1999; 99US-0159637.
PP 14-OCT-1999; 99US-0159638.
PP 18-OCT-1999; 99US-0159584.
PP 21-OCT-1999; 99US-0160741.
PP 21-OCT-1999; 99US-0160767.
PP 21-OCT-1999; 99US-0160768.
PP 21-OCT-1999; 99US-0160770.
PP 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PP 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PP 25-OCT-1999; 99US-0161399.
PR 25-OCT-1999; 99US-0161404.
PP 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161363.
PR 26-OCT-1999; 99US-0161364.
PR 27-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PP 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162141.

Query Match 98.2%; Score 49; DB 21; Length 448;
Post Local Similarity 98.2%; Pred. No. 90;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 WVPWHF 6
|| |||
Db 257 WVNWHF 262

RESULT 15
AAG47192
ID AAG47192 standard; Protein; 448 AA.
XX
AC AAG47192;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO 53453.
XX
KW Protein identification, signal transduction pathway; metabolic pathway;
KW hybridisation assay, genetic mapping; gene expression control, promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.

XX

PF 25-FEB-2000; 2000EP-0301439

XX

PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135829.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136792.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137602.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138034.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144630.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.

PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 29-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 88.9%; Score 40; DB 21; Length 448;
Best Local Similarity 83.3%; Pred. No. 90;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 WVRWHF 6
|| ||
Db 257 WVNWHF 262

Search completed: February 20, 2003, 10:47:34
Job time : 30.3333 secs

GenCore version 5.1.3

OM protein - protein search, using sw model

Run on: February 20, 2003, 10:48:13 ; Search Time 12 Seconds
(without alignments)
48.067 Million cell updates/sec

Title: US-09-912-414-2
Perfect score: 45
Sequence: 1 WVRWHF 6

Scoring table: BLOSUM62DX
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_71:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	40	88.9	443	2	T08136	probable omega-6 d
2	40	88.9	448	2	D85362	hypothetical prote
3	39	86.7	293	2	AB3299	hypothetical prote
4	39	86.7	370	2	A96741	hypothetical prote
5	39	86.7	482	2	S49465	catalase (EC 1.11.
6	38	84.4	477	1	I73631	brain-derived neur
7	38	84.4	822	1	A56853	brain-derived neur
8	37	82.2	591	2	G01586	probable protein 5
9	36	80.0	186	2	T44538	hypothetical prote
10	36	80.0	230	2	D70847	hypothetical prote
11	36	80.0	262	2	H83214	probable permease
12	36	80.0	265	2	AC0441	probable aliphatic
13	36	80.0	266	2	AD1238	hypothetical prote
14	36	80.0	384	2	E75295	conserved hypothet
15	36	80.0	424	2	T07742	omega-6 desaturase
16	36	80.0	472	2	AG2417	hypothetical prote
17	36	80.0	754	2	AB0614	probable competenc
18	36	80.0	1041	2	T29010	hypothetical prote
19	35	77.8	111	2	AB0690	probable membrane
20	35	77.8	226	1	F64450	hypothetical prote
21	35	77.8	458	2	B81409	probable transmemb
22	35	77.8	508	2	G90372	4 hydroxyphenylace
23	35	77.8	718	2	T29448	hypothetical prote
24	34	75.6	96	4	QQEC31	hypothetical prote
25	34	75.6	151	2	E64430	ribosomal protein
26	34	75.6	179	2	B83344	hypothetical prote
27	34	75.6	218	2	T27954	hypothetical prote
28	34	75.6	265	2	AC0433	conserved hypothet
29	34	75.6	301	2	G83352	transcription regu
30	34	75.6	316	2	C47099	lysR homolog, nac
31	34	75.6	316	2	C64963	transcription regu
32	34	75.6	316	2	E85823	transcription regu
33	34	75.6	316	2	G90976	transcription regu
34	34	75.6	329	2	AF1991	hypothetical prote
35	34	75.6	358	1	S43876	site-specific DNA
36	34	75.6	358	2	A87230	modification methy
37	34	75.6	394	2	D90185	putative SAF DNA-b
38	34	75.6	424	2	JC5891	omega 6 desaturase
39	34	75.6	495	2	S52273	active catalase /fla
40	34	75.6	497	2	T40586	nucleolar protein
41	34	75.6	522	2	[illegible]	nucleolar protein
42	34	75.6	559	2	G95104	hypothetical prote
43	34	75.6	709	2	T48989	serine/threonine p
44	34	75.6	910	2	C69456	subtilisin sendai
45	33.5	74.4	1511	2	S63932	probable membrane

ALIGNMENTS

RESULT 1
T08136
probable omega-6 desaturase (EC 1.14.99.-) precursor, chloroplast - rape
C;Species: Brassica napus (rape)
C;Date: 21-May-1999 #sequence_revision 21-May-1999 #text_change 08-Oct-1999
C;Accession: T08136
R;Hitz, W.D.; Carlson, T.J.; Booth, J.R.; Kinney, A.J.; Stecca, K.L.; Yadav, N.S.
Plant Physiol. 105, 635-641, 1994
A;Title: Cloning of a higher-plant plastid omega-6 fatty acid desaturase cDNA and its exp
A;Reference number: Z16109; MUID:94345008; PMID:8066133
A;Accession: T08136
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-443 <HIT>
A;Cross-references: EMBL:L29214, NID:g407930, PIDN:AAA33115.1, PID:g407931
A;Experimental source: seed
C;Genetics:
A;Genome: nuclear
C;Superfamily: omega-3 fatty acid desaturase
C;Keywords: chloroplast; fatty acid metabolism; oxidoreductase

Query Match 88.9%; Score 40; DB 2; Length 443;
Best Local Similarity 83.3%; Pred. No. 31;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy      1 WVRWHF 6
          || |||
Db    252 WVNWHF 257
```

RESULT 2
D85362
hypothetical protein AT4g30950 [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse ear cress)
C;Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 17-May-2002
C;Accession: D85362
R;anonymous; The European Union Arabidopsis Genome Sequencing Consortium, The Cold Spring
Nature 402, 769-777, 1999
A;Title: Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana.
A;Reference number: A85021; MUID:20083498; PMID:10617198
A;Accession: D85362
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-448 <STO>
A;Cross-references: GB:NC_003078, NID:g7573977, PIDN:AAF73913.1, GENP:NP_567110
C;Genetics:
A;Gene: AT4g30950
A;Map position: 4
C;Superfamily: omega-3 fatty acid desaturase

Query Match 88.9%; Score 40; DB 2; Length 448;
Best Local Similarity 83.3%; Pred. No. 31;

```
  Matches    5;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 WVRWHF 6
           |:|:||
Db     257 WVNWHF 262


RESULT 3
A83299
hypothetical protein PA2778 [imported] - Pseudomonas aeruginosa (strain PAO1)
C;Species: Pseudomonas aeruginosa
C;Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 21-Dec-2001
C;Accession: A83299
R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrener, P.; Hickey, M.J.; Br
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,
.; Lory, S.; Sisco, M.V.
Nature 406, 959-964, 2000
A;Title: Complete genome sequence of Pseudomonas aeruginosa PAO1, an opportunistic patho
A;Reference number: A82950; MUID:20403337; PMID:10984043
A;Accession: A83299
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1 293 <STO>
A;Cross-references: GB:AE004431; GB:AE004431; NID:g9949611; PIDN:AAG05130.1; GSPDB:GN00
A;Experimental source: strain PAO1
C;Genetics:
A;Gene: PA2778

  Query Match             86.7%;  Score 39;  DB 2;  Length 293;
  Best Local Similarity   83.3%;  Pred. No. 31;
  Matches    5;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 WVRWHF 6
           |:|:||
Db     101 WPRWHF 106


RESULT 4
A96741
hypothetical protein F14O23.22 [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C;Accession: A96741
R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
 Chin, C.W.; Chung, M.K.; Conn, T.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A;Authors: Hunter, C.; Tamkins, J.; Johnson-Hopson, C.; Khan, S.; Khazi, F.; Kim, J.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, G.X.; Liu, Z.A.; Loros, J.G.; Maiti, R.; Marshall,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis
A;Reference number: A86141; MUID:21016719; PMID:11130712
A;Accession: A96741
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-370 <STO>
A;Cross-references: GB:AF065173; NID:g7239511; PIDN:AAF43237.1; GSPDB:GN00141
C;Genetics:
A;Gene: F14O23.22
A;Map position: 1

  Query Match             86.7%;  Score 39;  DB 2;  Length 370;
  Best Local Similarity   100.0%;  Pred. No. 38;
  Matches    5;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 WVRWH 5
           |:|:|
Db     151 WVRWH 155


RESULT 5
S49465
catalase (EC 1.11.1.6) - nematode (Onchocerca volvulus)
C;Species: Onchocerca volvulus
C;Date: 01-Feb-1995 #sequence_revision 10-Nov-1995 #text_change 04-Mar-2000
C;Accession: S49465
R;Henkle-Duehrsen, K.J.
submitted to the EMBL Data Library, October 1994
A;Reference number: S49452
A;Accession: S49465
A;Molecule type: mRNA
A;Residues: 1-492 <HEN>
A;Cross-references: EMBL:X64156; NID:g19654; PIDN:CAA45661.1; PID:g28625
C;Superfamily: catalase
C;Keywords: chromoprotein, heme; iron; metal-protein; oxidoreductase
F;58,94,128/Active site: His, Ser, Asn #status predicted
F;338/Binding site: heme iron (Tyr) (axial ligand) #status predicted

  Query Match             86.7%;  Score 39;  DB 2;  Length 482;
  Best Local Similarity   83.3%;  Pred. No. 48;
  Matches    5;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 WVRWHF 6
           |:|:||
Db     216 WVRWHT 221


RESULT 6
I73631
brain-derived neurotrophic factor receptor precursor, short splice form - human
N;Alternate names: neurotrophin receptor trkB; truncated receptor tyrosine kinase trkB
C;Species: Homo sapiens (man)
C;Date: 02-Aug-1996 #sequence_revision 02-Aug-1996 #text_change 11-Jun-1999
C;Accession: I73631; I38357; S66384; S44129
R;Shelton, D.L.; Sutherland, J.; Gripp, J.; Camerato, T.; Armanini, M.P.; Phillips, H.S.;
J. Neurosci. 15, 477-491, 1995
A;Title: Human trks: molecular cloning, tissue distribution, and expression of extracell
A;Reference number: I56557; MUID:95123473; PMID:7823156
A;Accession: I73631
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-477 <RES>
A;Cross-references: GB:S76471; NID:g313719; PIDN:AAB33110.1; PID:g913720
R;Allen, S.J.; Dawbarn, D.; Eckford, S.D.; Wilcock, G.K.; Ashurst, M.; Scott-Brown, J.M.;
Neuroscience 60, 925-934, 1994
A;Title: Cloning of a non-catalytic form of human trkB and distribution of its mRNA
A;Reference number: I38357; MUID:95022162; PMID:7936202
A;Accession: I38357
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-477 <RE2>
A;Cross-references: EMBL:Z26174; NID:g476907; PIDN:CAA81971.1; PID:g47/3008
R;Hanin, M.; Talvenheimo, J.; Lu, J.; Futta, V.; Welcher, A.; Rohde, M.F.
Arch. Pl. Chem. Pl Phys. 25D, 259-264, 1995
A;Title: Extracellular domain of neurotrophin receptor trkB: disulfide structure, N-glyco
A;Reference number: S66384; MUID:96004804; PMID:7574684
A;Accession: S66384
A;Molecule type: protein
A;Residues: 1-32,33-42,43-57,143-156,158-165,181-199,302-323,333-353,354-357,'X',358-48
C;Comment: This form of the receptor is missing the protein kinase domain.
C;Genetics:
A;Gene: GDB:NTRK2; trkB
A;Cross-references: GDB:187993; OMIM:600456
A;Map position: 9q22.1-9q22.1
C;Superfamily: nerve growth factor receptor, high affinity, leucine rich alpha 2-glycopro
C;Keywords: alternative splicing; glycoprotein; growth factor receptor; tandem repeat; tr
F;1-31/Domain: signal sequence #status predicted <SIG>
F;32-477/Product: brain-derived neurotrophic factor receptor, short splice form #status p
F;32-430/Domain: extracellular #status predicted <EXT>
F;67-91/Domain: leucine rich alpha 2 glycoprotein repeat homology <LRR1>
F;92-115/Domain: leucine rich alpha 2 glycoprotein repeat homology <LRR2>
F;116-138/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR3>
F;139-159/Domain: leucine rich alpha-2-glycoprotein repeat homolog; #status predicted <LRR
F;430-452/Domain: transmembrane #status predicted <TMN>
```

```
F;453-477/Domain: cytosolic #status predicted <CYT>
F;32-38,36-49,152-176,194-194,219-266,302-345/Disulfide bonds: #status experimental
F;67,95,121,179,205,241,264,290,325,338,350,412/Binding site: carbohydrate (Asn) (covale

  Query Match             84.4%;  Score 38;  DB 1;  Length 477;
  Best Local Similarity   80.0%;  Pred. No. 69;
  Matches    4;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 WVRWH 5
            |:|||
Db       4 WIRWH 8


RESULT 7
A56853
brain-derived neurotrophic factor receptor precursor - human
N;Alternate names: receptor tyrosine kinase trkB
N;Contains: protein-tyrosine kinase (EC 2.7.1.112)
C;Species: Homo sapiens (man)
C;Date: 11-Aug-1995 #sequence_revision 11-Aug-1995 #text_change 11-Jun-1999
C;Accession: A56853; I56557
R;Nakagawara, A.; Liu, X.G.; Ikegaki, N.; White, P.S.; Yamashiro, D.J.; Nycum, L.M.; Bie
Genomics 25, 538-546, 1995
A;Title: Cloning and chromosomal localization of the human TPK-B tyrosine kinase recepto
A;Reference number: A56853; MUID:95309922; PMID:7789988
A;Accession: A56853
A;Molecule type: mRNA
A;Residues: 1-822 <NAK>
A;Cross-references: GB:U12140; NID:g535313; PIDN:AAC51471.1; PID:g538791
R;Shelton, D.L.; Sutherland, J.; Gripp, J.; Camerato, T.; Armanini, M.P.; Phillips, H.S.
J. Neurosci. 15, 477-491, 1995
A;Title: Human trks: molecular cloning, tissue distribution, and expression of extracell
A;Reference number: I56557; MUID:95123473; PMID:7823156
A;Accession: I56557
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-822 <SHE>
A;Cross-references: GB:S76473; NID:g913717; PIDN:AAB32100.1; PID:g913718
C;Genetics:
A;Gene: GDB:NTRK2; trkB
A;Cross-references: GDB:120899; OMIM:600456
A;Map position: 9q22.1-9q22.1
C;Function:
A;Description: regulation of nervous system development, receptor for brain derived neur
C;Superfamily: nerve growth factor receptor, high affinity, leucine-rich alpha 2-glycopr
C;Keywords: alternative splicing; ATP; autophosphorylation; glycoprotein; growth factor;
inase
F;1-31/Domain: signal sequence #status predicted <SIG>
F;32-822/Product: brain-derived neurotrophic factor receptor #status predicted <MAT>
F;32-438/Domain: extracellular #status predicted <EXT>
F;67-91/Domain: leucine rich alpha 2 glycoprotein repeat homology <LRR1>
F;92-115/Domain: leucine-rich alpha 2 glycoprotein repeat homology <LRR2>
F;116-138/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR3>
F;139-150/Domain: leucine rich alpha 2 glycoprotein repeat homology #status atypical <LR
F;436-452/Domain: transmembrane #status predicted <TMN>
F;453-822/Domain: cytosolic #status predicted <CYT>
F;536-814/Domain: protein kinase homology <KIN>
F;544-552/Region: protein kinase ATP binding motif
F;67,95,121,179,205,241,264,290,325,338,350,412/Binding site: carbohydrate (Asn) (covale
F;572/Active site: Lys #status predicted
F;706/Binding site: phosphate (Tyr) (covalent) (by autophosphorylation) #status predicte
F;817/Binding site: phosphate (Tyr) (covalent) #status predicted

  Query Match             84.4%;  Score 38;  DB 1;  Length 822;
  Best Local Similarity   80.0%;  Pred. No. 1.1e+02;
  Matches    4;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 WVRWH 5
            |:|||
Db       4 WIRWH 8


RESULT 8
G01586
probable protein 55.11 homolog p67 [similarity] - human
C;Species: Homo sapiens (man)
C;Date: 21-Dec-1996 #sequence_revision 06-Jun-1997 #text_change 20-Jun-2000
C;Accession: G01586
R;Yew, F.
submitted to the EMBL Data Library, December 1994
A;Reference number: G07854
A;Accession: G01586
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-591 <YEW>
A;Cross-references: EMBL:U18247; NID:g603224; PIDN:AAA57339.1; PID:g603225

  Query Match             82.2%;  Score 37;  DB 2;  Length 591;
  Best Local Similarity   66.7%;  Pred. No. 1.2e+02;
  Matches    4;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 WVRWHF 6
            | |||:
Db     350 WCRWHY 355


RESULT 9
T44538
hypothetical protein VP2 [imported] - Pseudomonas aeruginosa
C;Species: Pseudomonas aeruginosa
C;Date: 31-Jan-2000 #sequence_revision 31-Jan-2000 #text_change 06-Oct-2000
C;Accession: T44538; B83570
R;Nakayama, K.; Takashima, K.; Ishihara, H.; Shinomiya, T.; Kageyama, M.; Kanaya, S.; Oh
submitted to the EMBL Data Library, August 1999
A;Description: Genetic relationship between bacteriocins and bacteriophages.
A;Reference number: S22790
A;Accession: T44538
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-185 <NAK>
A;Cross-references: EMBL:AB030805; PIDN:BAA83153.1
A;Experimental source: strain PAO1
R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrener, P.; Hickey, M.J.; Bri
adman, F.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A;Title: Complete genome sequence of Pseudomonas aeruginosa PAO1, an opportunistic pathog
A;Reference number: A82950; MUID:20437337; PMID:10984043
A;Accession: B83570
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-185 <STO>
A;Cross-references: GB:AE004461; GB:AE004461; NID:g9947[illegible]; PIDN:AAG04[illegible].1; GSPDB:GN00[illegible]
A;Experimental source: strain PAO1
C;Genetics:
A;Gene: PA0616
C;Superfamily: Pseudomonas aeruginosa hypothetical protein VP2

  Query Match             80.0%;  Score 36;  DB 2;  Length 185;
  Best Local Similarity   80.0%;  Pred. No. 60;
  Matches    4;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 WVRWH 5
            |:|||
Db      42 WLRWH 46


RESULT 10
D70847
hypothetical protein Rv0059 - Mycobacterium tuberculosis (strain H37RV)
C;Species: Mycobacterium tuberculosis
C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C;Accession: D70847
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.
, Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.;
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
```

Nature 393, 537 S14, 1998
A;Authors: Soares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A;Reference number: AC0000; MUID:98295987; PMID:9634230
A;Accession: AC0947
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-230 <COL>
A;Cross-references: GB:AL009190; GP:AL123456; NID:g2621614; PIDN:CAA18240.1; PID:g2621
A;Experimental source: strain H37Rv
C;Genetics:
A;Gene: Rv0059
C;Superfamily: Mycobacterium tuberculosis hypothetical protein Rv0059

  Query Match 80.0%; Score 36; DB 2; Length 230;
  Best Local Similarity 80.0%; Pred. No. 53;
  Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 WVRWH 5
|:||
Db 26 WIVWHF 31

RESULT 11
H83214
probable permease of ABC transporter PA3443 [imported] - Pseudomonas aeruginosa (strain
C;Species: Pseudomonas aeruginosa
C;Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C;Accession: H83214
R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrener, P.; Hickey, M.J.; Br
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A;Title: Complete genome sequence of Pseudomonas aeruginosa PAO1, an opportunistic patho
A;Reference number: A59477; MUID:20430418; PMID:10984043
A;Accession: H83214
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-262 <STO>
A;Cross-references: GB:AE004765; GB:AE004091; NID:g9949580; PIDN:AAG06831.1; GSPDB:GN001
A;Experimental source: strain PAO1
C;Genetics:
A;Gene: PA3443
C;Superfamily: Synechococcus nitrate transport protein nrtB

  Query Match 80.0%; Score 36; DB 2; Length 262;
  Best Local Similarity 80.0%; Pred. No. 90;
  Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 WVRWH 5
|:||
Db 248 WLRWH 252

RESULT 12
AC0441
probable aliphatic sulfonates transport permease protein ssuC [imported] - Yersinia pest
C;Species: Yersinia pestis
C;Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 02-Nov-2001
C;Accession: AC0441
R;Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.
dem, Turnbull, P.C.; Fraser, C.M.; Dougan, G.
Nature 413, 523-527, 2001
A;Title: Genome sequence of Yersinia pestis, the causative agent of plague.
A;Reference number: AB0001; MUID:21470413; PMID:11586360
A;Accession: AC0441
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-265 <PUB>
A;Cross-references: GB:AL590842; PIDN:CAC92615.1; PID:g16122684; GSPDB:GN00175
C;Genetics:
A;Gene: ssuC
C;Superfamily: Synechococcus nitrate transport protein nrtB

  Query Match 80.0%; Score 36; DB 2; Length 265;
  Best Local Similarity 80.0%; Pred. No. 83;
  Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 WVRWH 5
|:||
Db 250 WLRWH 254

RESULT 13
AD3238
hypothetical protein nrtB [imported] - Agrobacterium tumefaciens (strain C58, Dupont) pl
C;Species: Agrobacterium tumefaciens
C;Date: 31-Jan-2001 #sequence_revision 31-Jan-2001 #text_change 01-Feb-2002
C;Accession: AD3238
R;Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.B.; Chan, Y.; Woo, L
drage, G.; Gillet, W.; Brunt, C.; Quetchner, D.; Matyasik, T.; Levy, R.; Li, M.; McClell
; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A;Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm, B
ster, E.W.
A;Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A;Reference number: AB2577; PMID:11743193
A;Accession: AD3238
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-266 <PUR>
A;Cross-references: GB:AE008688; PIDN:AAL46822.1; PID:g17744108; GSPDB:GN00189
A;Experimental source: strain C58 (Dupont)
C;Genetics:
A;Gene: nrtB
A;Genome: plasmid
C;Superfamily: Synechococcus nitrate transport protein nrtB

  Query Match 80.0%; Score 36; DB 2; Length 266;
  Best Local Similarity 80.0%; Pred. No. 83;
  Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 WVRWH 5
|:||
Db 257 WLRWH 261

RESULT 14
E75295
hypothetical protein - Deinococcus radiodurans (strain R1)
C;Species: Deinococcus radiodurans
C;Date: 05-Dec-1999 #sequence_revision 05-Dec-1999 #text_change 17-Mar-2000
C;Accession: E75295
R;White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.; H
, M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Kar
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A;Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A;Reference number: A75295; MUID:20001866; PMID:10567266
A;Accession: E75295
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-284 <WHI>
A;Cross-references: GB:AE000513; GB:AE000513; NID:g4461783; PIDN:AAF11801.1; PID:g4460
A;Experimental source: strain R1
C;Genetics:
A;Gene: DR2253
A;Map position: 1

  Query Match 80.0%; Score 36; DB 2; Length 384;
  Best Local Similarity 80.0%; Pred. No. 1.2e+02;
  Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 WVRWH 5
|:||

```
Db    362 WLRWH 366


RESULT 15
T07742
omega-6 desaturase, chloroplast - soybean
C;Species: Glycine max (soybean)
C;Date: 14-May-1999 #sequence_revision 14-May-1999 #text_change 08-Oct-1999
C;Accession: T07742
R;Hitz, W.D ; Carlson, T.J ; Booth, J.R ; Kinney, A.J ; Stecca, K.L ; Yadav, N.S.
Plant Physiol. 105, 635-641, 1994
A;Title: Cloning of a higher-plant plastid omega-6 fatty acid desaturase cDNA and its ex
A;Reference number: Z16109; MUID:94345008; PMID:8066133
A;Accession: T07742
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-424 <HIT>
A;Cross-references: EMBL:L29215; NID.g459961; PIDN:AAA50158.1; PID:g459962
A;Experimental source: seed
C;Genetics:
A;Genome: nuclear
C;Superfamily: omega-3 fatty acid desaturase
C;Keywords: chloroplast

  Query Match             80.0%;  Score 36;  DB 2;  Length 424;
  Best Local Similarity   66.7%;  Pred. No. 1.3e+02;
  Matches    4;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 WVRWHF 6
          |: |||
Db    251 WLMWHF 256


Search completed: February 20, 2003, 10:50:00
Job time : 13 secs
```

GenCore version 5.1.3

OM protein - protein search, using sw model

Run on: February 20, 2003, 10:41:55 , Search time 6.66667 Seconds
(without alignments)
37 329 Million cell updates/sec

Title: US-09-912-414-2
Perfect score: 45
Sequence: 1 WVRWHF 6

Scoring table: BLOSUM62DX
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	42	93.3	479	1	CATA_PSEPU	Q59714 pseudomonas
2	40	88.9	443	1	FD6C_BRANA	P48627 brassica na
3	40	88.9	448	1	FD6C_ARATH	P46312 arabidopsis
4	39	86.7	482	1	CATA_ONCVE	Q27710 onchocerca
5	38	84.4	822	1	TPKR_HUMAN	Q16620 homo sapien
6	36	80.0	474	1	FD6C_SOYBN	P48628 glycine max
7	36	80.0	788	1	BCSB_XANAC	P58933 xanthomonas
8	36	80.0	1169	1	SUV3_DROME	P20193 drosophila
9	35	77.8	226	1	YC07_METJA	Q58604 methanococc
10	35	77.8	718	1	PLSB_CAERE	Q22949 caenorhabdi
11	34	75.6	148	1	RS12_METJA	P54062 methanococc
12	34	75.6	316	1	CBL_ECOLI	Q47083 escherichia
13	34	75.6	316	1	CBL_KLEAE	Q08598 klebsiella
14	34	75.6	358	1	MTC1_CAUCR	Q45971 caulobacter
15	34	75.6	372	1	GP45_HUMAN	Q9y5y3 homo sapien
16	34	75.6	373	1	GP45_MOUSE	Q9eqq4 mus musculu
17	34	75.6	495	1	AOFN_ASPNG	P46882 aspergillus
18	34	75.6	799	1	AFSK_STPCO	P54741 streptomyce
19	34	75.6	807	1	AFSK_STPGR	P54742 streptomyce
20	34	75.6	1698	1	Y076_HUMAN	Q14999 homo sapien
21	33.5	74.4	1511	1	PDRC_YEAST	Q02785 saccharomyc
22	33	73.3	244	1	YU68_CAERE	P52057 caenorhabdi
23	33	73.3	376	1	MTS1_PHIME	O30569 rhizobium m
24	33	73.3	377	1	MTB1_BRUAB	O30570 brucella ab
25	33	73.3	396	1	BCP_ECOLI	P28246 escherichia
26	33	73.3	439	1	Y412_ARATH	O04658 arabidopsis
27	33	73.3	478	1	CATA_LACSK	P30265 lactobacill
28	33	73.3	483	1	BCA_STRVL	P33569 streptomyce
29	33	73.3	497	1	CATA_MICLU	P29422 micrococcus
30	33	73.3	502	1	CATA_TOXGO	Q9xzd5 toxoplasma
31	33	73.3	504	1	SIK1_YEAST	Q12460 saccharomyc
32	33	73.3	511	1	NOP5_YEAST	Q12499 saccharomyc
33	33	73.3	529	1	NOP5_HUMAN	Q9y2x3 homo sapien
34	33	73.3	534	1	NOP5_RAT	Q9qz86 rattus norv
35	33	73.3	614	1	YG3F_YEAST	P53283 saccharomyc
36	33	73.3	821	1	TPKR_MOUSE	P15209 mus musculu
37	33	73.3	821	1	TPKR_RAT	Q63604 rattus norv
38	33	73.3	1015	1	FDNG_ECOLI	P24183 escherichia
39	33	73.3	1016	1	FDOG_ECOLI	P32176 escherichia
40	32	71.1	129	1	YHW2_YEAST	P38857 saccharomyc
41	32	71.1	146	1	YG56_YEAST	P53311 saccharomyc
42	32	71.1	287	1	YAHE_ECOLI	P77207 escherichia
43	32	71.1	386	1	RL4_UREPA	P49165 urechis cau
44	32	71.1	395	1	DLTB_BACSU	P39580 bacillus su
45	32	71.1	396	1	RL4A_XENLA	P08427 xenopus lae

ALIGNMENTS

RESULT 1
CATA_PSEPU
ID CATA_PSEPU STANDARD; PRT; 479 AA.
AC Q59714;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Catalase (EC 1.11.1.6).
GN KATA OR CATA.
OS Pseudomonas putida.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=303;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Corvallis;
RX MEDLINE=98010091; PubMed=9358059;
RA Kim Y.C., Miller C.D., Anderson A.J.;
RT "Identification of adjacent genes encoding the major catalase and a
RT bacterioferritin from the plant-beneficial bacterium Pseudomonas
RT putida.";
RL Gene 199:219-224(1997).
CC -!- FUNCTION: DECOMPOSES HYDROGEN PEROXIDE IN WATER AND OXYGEN, SERVES
CC TO PROTECT CELLS FROM THE TOXIC EFFECTS OF HYDROGEN PEROXIDE.
CC -!- CATALYTIC ACTIVITY: 2 H(2)O(2) = O(2) + 2 H(2)O.
CC -!- COFACTOR: HEME GROUP.
CC -!- ENZYME REGULATION: ACTIVATED BY PEROXIDE.
CC -!- SIMILARITY: BELONGS TO THE CATALASE FAMILY.
CC --
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC --
DR EMBL; U63511; AAB88219.1; -.
DR HSSP; P42321; 2CAE.
DR InterPro; IPR002226; Catalase.
DR Pfam; PF00199; catalase; 1.
DR PRINTS; PR00067; CATALASE.
DR ProDom; PD000510; Catalase; 1.
DR PROSITE; PS00437; CATALASE_1; 1.
DR PROSITE; PS00438; CATALASE_2; 1.
KW Oxidoreductase; Peroxidase; Iron; Heme; Hydrogen peroxide.
FT ACT_SITE 53 53 BY SIMILARITY.
FT ACT_SITE 126 126 BY SIMILARITY.
FT BINDING 336 336 PROXIMAL HEME LIGAND (BY SIMILARITY).
SQ SEQUENCE 479 AA; 53381 MW; FFE3CBDB67778571 CRC64;

Query Match 93.3%, Score 42, DB 1, Length 479;
Best Local Similarity 83.3%; Pred. No. 3.4;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 WVRWHF 6

```
Db    209 WVKWHF 914


RESULT 2
FD6C_BRANA
ID   FD6C_BRANA     STANDARD;      PRT;   443 AA.
AC   P48627;
DT   01-FEB-1996 (Rel. 33, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Omega-6 fatty acid desaturase, chloroplast precursor (EC 1.14.99.-).
OS   Brassica napus (Rape).
OC   Eukaryota, Viridiplantae, Streptophyta, Embryophyta, Tracheophyta;
OC   Spermatophyta, Magnoliophyta, eudicotyledons, core eudicots, Rosidae;
OC   eurosids II; Brassicales, Brassicaceae, Brassiceae.
OX   NCBI_TaxID=3708;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Seed;
RX   MEDLINE=94345[illegible], PubMed=[illegible];
RA   Hitz W.D., Carlson T.J., Booth J.R. Jr., Kinney A.J., Shanklin J.;
RA   Yadav N.S.;
RT   "Cloning of a higher plant plastid omega-6 fatty acid desaturase cDNA
RT   and its expression in a [illegible] yeast.";
RL   Plant Physiol. 105:635-641(1994);
CC   -!- FUNCTION: CHLOROPLAST OMEGA-6 FATTY ACID DESATURASE (MICROSOMAL
CC       THE ENZYME PARTICIPATES IN THE BIOSYNTHESIS OF 16:3 AND 18:3 FATTY
CC       ACIDS, IMPORTANT CONSTITUENTS OF PLANT MEMBRANES. IT IS THOUGHT
CC       TO USE FERREDOXIN AS AN ELECTRON DONOR AND TO ACT ON FATTY ACIDS
CC       ESTERIFIED TO GALACTOLIPIDS, SULFOLIPIDS AND PHOSPHATIDYLGLYCEROL.
CC   -!- PATHWAY: Polyunsaturated fatty acid biosynthesis.
CC   -!- SUBCELLULAR LOCATION: CHLOROPLAST, MEMBRANE BOUND (PROBABLE).
CC   -!- DOMAIN: THE HISTIDINE BOX DOMAINS MAY CONTAIN THE ACTIVE SITE
CC       AND/ OR BE INVOLVED IN METAL ION BINDING.
CC   -!- SIMILARITY: BELONGS TO THE FATTY ACID DESATURASE FAMILY.
CC   ----------------------------------------------------------------------
CC   This [illegible] copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to licenser@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; L29214; AAA33157.1; -.
DR   InterPro; IPR001225; FA_desaturase.
DR   Pfam; PF00487; FA_desaturase; 1.
DR   ProDom; PD001081; FA_desaturase; 1.
KW   Oxidoreductase; Fatty acid biosynthesis; Chloroplast; Membrane;
KW   Transit peptide.
FT   TRANSIT       1     65       CHLOROPLAST (BY SIMILARITY).
FT   CHAIN        65    443       OMEGA-6 FATTY ACID DESATURASE
FT   DOMAIN      166    170       HISTIDINE BOX 1.
FT   DOMAIN      202    206       HISTIDINE BOX 2.
FT   DOMAIN      362    366       HISTIDINE BOX 3.
SQ   SEQUENCE   443 AA;  [illegible] MW;  [illegible] CRC64;

  Query Match             88.3%;  Score 40;  DB 1;  Length 443;
  Best Local Similarity   93.8%;  Pred. No. [illegible];
  Matches   5;  Conservative   0;  Mismatches   1;  Indels   0;  Gaps   0;

Qy      1 WVRWHF 6
          || |||
Db    252 WVNWHF 257


RESULT 3
FD6C_ARATH
ID   FD6C_ARATH     STANDARD;      PRT;   449 AA.
AC   P46312, Q9M034;
DT   01-NOV-1995 (Rel. 32, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Omega-6 fatty acid desaturase, chloroplast precursor (EC 1.14.99.-).
GN   FAD6 OR FAD7 OR AI4010850 OR F6I18.140.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota, Viridiplantae, Streptophyta, Embryophyta, Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE OF 1-419 FROM N.A.
RC   STRAIN=cv. Columbia;
RX   MEDLINE=[illegible], PubMed=8846158;
RA   Falcone D.L., Gibson S., Lemieux B., Somerville C.R.;
RT   "Identification of a gene that complements an Arabidopsis mutant
RT   deficient in chloroplast omega-6 desaturase activity.";
RL   Plant Physiol. 106:1453-1459(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Columbia;
RX   MEDLINE=[illegible], PubMed=10617198;
RA   Mayer K.F.X., Schoof H., Haecker A., Lenhard M., Weitkamp J.,
RA   Hall T., Chesterfield A., Stiekema W., Berlan R. D., Terryn N.,
RA   Harris B., Ansorge W., Brandt P., Grivell L., Rieger M.,
RA   Weichselgartner M., de Simone V., Obermaier B., Mache R., Mueller M.,
RA   Kreis M., Delseny M., Puigdomenech P., Watson M., Schmidtheini T.,
RA   Reichert B., Portetelle D., Perez-Alonso M., Boutry M., Bancroft I.,
RA   Vos P., Hoheisel J., Zimmermann W., Wedler H., Ridley P.,
RA   Langham S. A., McCullagh B., Bilham L., Robben J.,
RA   Van der Schueren J., Grymonprez B., Chuang Y. J., Vandenbussche F.,
RA   Braeken M., Weltjens I., Voet M., Bastiaens I., Aert R., Defoor E.,
RA   Weitzenegger T., Bothe G., Ramsperger U., Hilbert H., Braun M.,
RA   Holzer E., Brandt A., Peters S., van Staveren M., Dirkse W.,
RA   Mooijman P., Klein Lankhorst R., Rose M., Hauf J., Koetter P.,
RA   Berneiser H., Hempel S., Feldpausch M., Lamberth S., Van den Daele H.,
RA   De Keyser A., Buysshaert C., Gielen J., Villarroel R., De Clercq R.,
RA   Van Montagu M., Rogers J., Cronin A., Quail M., Bray-Allen S.,
RA   Clark L., Doggett J., Hall S., Kay M., Lennard N., McLay K., Mayes R.,
RA   Pettett A., Rajandream M.A., Lyne M., Benes V., Rechmann S.,
RA   Borkova D., Blocker H., Scharfe M., Grimm M., Loehnert T.-H.,
RA   Dose S., de Haan M., Maarse A., Schaefer M., Mueller-Auer S.,
RA   Gabel C., Fuchs M., Fartmann B., Granderath K., Dauner D., Herzl A.,
RA   Neumann S., Argiriou A., Vitale D., Liguori R., Piravandi E.,
RA   Massenet O., Quigley F., Clabauld G., Muendlein A., Felber R.,
RA   Schnabl S., Hiller R., Schmidt W., Lecharny A., Aubourg S.,
RA   Chefdor F., Cooke R., Berger C., Monfort A., Casacuberta E.,
RA   Gibbons T., Weber N., Vandenbol M., Bargues M., Terol J., Torres A.,
RA   Perez-Perez A., Purnelle B., Bent E., Johnson S., Tacon D., Jesse T.,
RA   Heijnen L., Schwarz S., Scholler P., Heber S., Francs P., Bielke C.,
RA   Frishman D., Haase D., Lemcke K., Mewes H. W., Stocker S.,
RA   Zaccaria P., Bevan M., Wilson R.K., de la Bastide M., Habermann K.,
RA   Parnell L., Dedhia N., Gnoj L., Schutz K., Huang E., Spiegel L.,
RA   Sekhon M., Murray J., Sheet P., Cordes M., Abu-Threideh J.,
RA   Stoneking T., Kalicki J., Graves T., Harmon G., Edwards J.,
RA   Latreille P., Courtney L., Cloud J., Abbott A., Scott K., Johnson D.,
RA   Minx P., Bentley D., Fulton B., Miller N., Greco T., Kemp K.,
RA   Kramer J., Fulton L., Mardis E., Dante M., Pepin K., Hillier L.,
RA   Nelson J., Spieth J., Ryan E., Andrews S., Geletka C., Layman D.,
RA   Du H., Ali J., Berghoff A., Jones K., Drone K., Cotton M., Joshu C.,
RA   Antonoiu B., Zidanic M., Strong C., Sun H., Lamar B., Yordan C.,
RA   Ma P., Zhong J., Preston R., Vil D., Shekher M., Matero A., Shah R.,
RA   Swaby I.K., O'Shaughnessy A., Rodriguez M., Hoffman J., Till S.,
RA   Granat S., Shohdy N., Hasegawa A., Hameed A., Lodhi M., Johnson A.,
RA   Chen E., Marra M., Martienssen R., McCombie W.R.;
RT   "Sequence and analysis of chromosome 4 of the plant Arabidopsis
RT   thaliana.";
RL   Nature 402:769-777(1999).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Columbia;
RA   Shinozaki K., Davis R.W., Ecker J.R., Theologis A.;
RT   "RIKEN Arabidopsis full length cDNA clones (RAFLs) sequenced by the
RT   OSP consortium (Salk/Stanford/PGFC).";
RL   Submitted (SEP-2000) to the EMBL/GenBank/DDBJ databases.
```

```
CC   -!- FUNCTION: CHLOROPLAST OMEGA-6 FATTY ACID DESATURASE INTRODUCES
CC       THE SECOND DOUBLE BOND IN THE BIOSYNTHESIS OF 16:3 AND 18:3 FATTY
CC       ACIDS, IMPORTANT CONSTITUENTS OF PLANT MEMBRANES. IT IS THOUGHT
CC       TO USE FERREDOXIN AS AN ELECTRON DONOR AND TO ACT ON FATTY ACIDS
CC       ESTERIFIED TO GALACTOLIPIDS, SULFOLIPIDS AND PHOSPHATIDYLGLYCEROL.
CC   -!- PATHWAY: Polyunsaturated fatty acid biosynthesis.
CC   -!- SUBCELLULAR LOCATION: CHLOROPLAST, MEMBRANE-BOUND (PROBABLE).
CC   -!- DEVELOPMENTAL STAGE: HIGHEST LEVELS FOUND IN EXPANDING LEAVES.
CC   -!- DOMAIN: THE HISTIDINE BOX DOMAINS MAY CONTAIN THE ACTIVE SITE
CC       AND/OR BE INVOLVED IN METAL ION BINDING.
CC   -!- SIMILARITY: BELONGS TO THE FATTY ACID DESATURASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U09503; AAA92800.1; -.
DR   EMBL; AL022198; CAA18198.1; -.
DR   EMBL; AL161578; CAB79813.1; -.
DR   EMBL; AY045621; AAK73979.1; -.
DR   EMBL; AY058078; AAL24186.1; -.
DR   EMBL; AY058852; AAL24240.1; -.
DR   InterPro; IPR001225; FA_desaturase.
DR   Pfam; PF00487; FA_desaturase; 1.
DR   ProDom; PD001081; FA_desaturase; 1.
KW   Oxidoreductase; Fatty acid biosynthesis; Chloroplast; Membrane;
KW   Transit peptide.
FT   TRANSIT       1     69       CHLOROPLAST (BY SIMILARITY).
FT   CHAIN        70    448       OMEGA-6 FATTY ACID DESATURASE.
FT   DOMAIN      171    175       HISTIDINE BOX-1.
FT   DOMAIN      207    211       HISTIDINE BOX-2.
FT   DOMAIN      367    371       HISTIDINE BOX-3.
SQ   SEQUENCE   448 AA;  51225 MW;  03A07FB2FBF2E7 CRC64;

  Query Match             89.0%;  Score 40;  DB 1;  Length 448;
  Best Local Similarity   92.3%;  Pred. No. 6.3;
  Matches    5;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 WVRWHF 6
           || |||
Db     257 WVNWHF 262


RESULT 4
CATA_ONCVE
ID   CATA_ONCVE     STANDARD;      PRT;   482 AA.
AC   Q27710; O85499;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   Catalase (EC 1.11.1.6).
GN   CAT.
OS   Onchocerca volvulus endobacterium
OC   Bacteria; Proteobacteria, alpha subdivision; Rickettsiales;
OC   Rickettsiaceae; Wolbachieae
OX   NCBI_TaxID=77551;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Henkle-Duehrsen K.J.;
RL   Submitted (OCT-1994) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Henkle-Duehrsen K.J., Eckelt V.H.O., Wildenburg G., Blaxter M.,
RA   Walter R.D.;
RT   "Gene structure, activity and localization of a catalase from
RT   intracellular bacteria in Onchocerca volvulus.";
RL   Submitted (MAY-1998) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: DECOMPOSES HYDROGEN PEROXIDE IN WATER AND OXYGEN, SERVES
CC       TO PROTECT CELLS FROM THE TOXIC EFFECTS OF HYDROGEN PEROXIDE.
CC   -!- CATALYTIC ACTIVITY: 2 H(2)O(2) = O(2) + 2 H(2)O.
CC   -!- COFACTOR: HEME GROUP.
CC   -!- SIMILARITY: BELONGS TO THE CATALASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X82176; CAA57666.1; -.
DR   EMBL; AF069070; AAC79431.1; -.
DR   HSSP; P42321; 2CAE.
DR   InterPro; IPR002226; Catalase.
DR   Pfam; PF00199; catalase; 1.
DR   PRINTS; PR00067; CATALASE.
DR   ProDom; PD000510; Catalase; 1.
DR   PROSITE; PS00437; CATALASE_1; 1.
DR   PROSITE; PS00438; CATALASE_2; 1.
KW   Oxidoreductase; Peroxidase; Iron; Heme; Hydrogen peroxide.
FT   ACT_SITE     55     55       BY SIMILARITY.
FT   ACT_SITE    128    128       BY SIMILARITY.
FT   BINDING     338    338       PROXIMAL HEME LIGAND (BY SIMILARITY).
SQ   SEQUENCE   482 AA;  53709 MW;  B1DA5F49BA39F776 CRC64;

  Query Match             86.7%;  Score 39;  DB 1;  Length 482;
  Best Local Similarity   66.7%;  Pred. No. 11;
  Matches    4;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 WVRWHF 6
             ||:||:
Db     211 WVKWHY 216


RESULT 5
TRKB_HUMAN
ID   TRKB_HUMAN     STANDARD;      PRT;   822 AA.
AC   Q16620; Q16675;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   BDNF/NT-3 growth factors receptor precursor (EC 2.7.1.112) (TrkB
DE   tyrosine kinase) (GP145-TrkB) (Trk-B).
GN   NTRK2 OR TRKB.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A. (LONG FORM).
RC   TISSUE=Hippocampus;
RX   MEDLINE=95303922; PubMed=7789988;
RA   Nakagawara A., Liu X.-G., Ikegaki N., White P.S., Yamashiro D.J.,
RA   Nycum L.M., Biegel J.A., Brodeur G.M.;
RT   "Cloning and chromosomal localization of the human TRK-B tyrosine
RT   kinase receptor gene (NTRK2).";
RL   Genomics 25:538-546(1995).
RN   [2]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE (LONG AND TRUNCATED FORMS).
RC   TISSUE=Brain;
RX   MEDLINE=95123473; PubMed=7823156;
RA   Shelton D.L., Sutherland J., Gripp J., Camerato T., Armanini M.P.,
RA   Phillips H.S., Carroll K., Spencer S.D., Levinson A.D.;
RT   "Human trks: molecular cloning, tissue distribution, and expression
RT   of extracellular domain immunoadhesins.";
RL   J. Neurosci. 15:477-491(1995).
RN   [3]
RP   SEQUENCE FROM N.A. (TRUNCATED FORMS).
RC   TISSUE=Hippocampus;
RX   MEDLINE=95322162; PubMed=7636202;
RA   Allen S.J., Dawbarn D., Eckford S.D., Wilcock G.K., Ashcroft M.,
```

```
RA   Colebrook S.M., Feeney P., Maggiwan S.H.,
RT   "Cloning of a non catalytic form of human trkB and identification of
RT   messenger RNA for trkB in human brain.",
RL   Neuroscience 60:825-834(1994).
CC    !- FUNCTION: RECEPTOR FOR BRAIN-DERIVED NEUROTROPHIC FACTOR (BDNF),
CC       NEUROTROPHIN 3 AND NEUROTROPHIN 4/5 BUT NOT NERVE GROWTH FACTOR
CC       (NGF). INVOLVED IN THE DEVELOPMENT AND/OR MAINTENANCE OF THE
CC       NERVOUS SYSTEM. THIS IS A TYROSINE PROTEIN KINASE RECEPTOR. KNOWN
CC       SUBSTRATES FOR THE TRK RECEPTORS ARE SHC, PI 3 KINASE, AND PLC
CC       GAMMA 1.
CC    -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC       tyrosine phosphate.
CC    -!- SUBUNIT: EXISTS IN A DYNAMIC EQUILIBRIUM BETWEEN MONOMERIC (LOW
CC       AFFINITY) AND DIMERIC (HIGH AFFINITY) STRUCTURES (BY SIMILARITY).
CC     !- SUBCELLULAR LOCATION: Type I membrane protein.
CC    -!- ALTERNATIVE PRODUCTS: A TRUNCATED NON CATALYTIC FORM WHICH
CC       FUNCTION REMAINS STILL UNKNOWN IS PRODUCED BY ALTERNATIVE
CC       SPLICING. THE RATIO OF FULL-LENGTH TO TRUNCATED FORM IS HIGHER IN
CC       FETAL THAN IN ADULT BRAIN.
CC    -!- TISSUE SPECIFICITY: WIDELY EXPRESSED, MAINLY IN THE NERVOUS
CC       TISSUE. IN THE CNS, EXPRESSION IS OBSERVED IN THE CEREBRAL CORTEX,
CC       HIPPOCAMPUS, THALAMUS, PURKIJE FIBERS, GRANULAR LAYER OF THE
CC       CEREBELLUM, BRAIN STEM, AND SPINAL CORD. IN THE PERIPHERAL NERVOUS
CC       SYSTEM, IT IS EXPRESSED IN MANY CRANIAL GANGLIA, THE OPHTALMIC
CC       NERVE, THE VESTIBULAR SYSTEM, MULTIPLE FACIAL STRUCTURES, THE
CC       SUBMAXILLARY GLANDS, AND TYPICAL ROOT GANGLIA.
CC    -!- PTM: LIGAND MEDIATED AUTO PHOSPHORYLATION.
CC    -!- SIMILARITY: BELONGS TO THE TYR FAMILY OF PROTEIN KINASES, INSULIN
CC       RECEPTOR SUBFAMILY.
CC    -!- SIMILARITY: CONTAINS 2 LEUCINE RICH REPEATS (LRR).
CC    -!- SIMILARITY: CONTAINS 2 IMMUNOGLOBULIN LIKE C2 TYPE DOMAINS.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; U12140; AAC51371.1; -.
DR   EMBL; S76474; AAB33109.1; -.
DR   EMBL; S76474; AAB33110.1; -.
DR   EMBL; X75958; CAA53571.1; -.
DR   HSSP; P06213; 1IRK.
DR   Genew; HGNC:8032; NTRK2.
DR   MIM; 600456; -.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR001611; LRR.
DR   InterPro; IPR000483; LRR_Cterm.
DR   InterPro; IPR000372; LRR_Nterm.
DR   InterPro; IPR002011; RTKinaseII.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   Pfam; PF00047; ig; 1.
DR   Pfam; PF00069; pkinase; 1.
DR   Pfam; PF00560; LRR; 1.
DR   Pfam; PF01462; LRRNT; 1.
DR   Pfam; PF01463; LRRCT; 1.
DR   PRINTS; PR00109; TYRKINASE.
DR   ProDom; PD000001; Euk_pkinase; 1.
DR   SMART; SM00408; IGc2; 1.
DR   SMART; SM00082; LRRCT; 1.
DR   SMART; SM00013; LRRNT; 1.
DR   SMART; SM00219; TyrKc; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR   PROSITE; PS00239; RECEPTOR_TYR_KIN_II; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW   Transferase, Tyrosine-protein kinase, Transmembrane, ATP-binding,
KW   Phosphoprotein, Receptor, Glycoprotein, Neurogenesis, Signal,
KW   Leucine-rich repeat, Repeat, Immunoglobulin domain,
FW   Alternative splicing; Polymorphism.
FT   SIGNAL        1     31
FT   CHAIN        32    822       BDNF/NT-3 GROWTH FACTORS RECEPTOR.
FT   DOMAIN       32    430       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     431   454       POTENTIAL.
FT   DOMAIN       455   822       CYTOPLASMIC (POTENTIAL).
FT   REPEAT        72    93       LRR 1.
FT   REPEAT        96    117      LRR 2.
FT   DOMAIN       214   299       IG-LIKE C2-TYPE DOMAIN 1.
FT   DOMAIN       301   365       IG-LIKE C2-TYPE DOMAIN 2.
FT   DOMAIN       528   807       PROTEIN KINASE.
FT   NP_BIND      544   552       ATP (BY SIMILARITY).
FT   BINDING      572   572       ATP (BY SIMILARITY).
FT   ACT_SITE     676   676       BY SIMILARITY.
FT   MOD_RES      708   708       PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT   MOD_RES      712   712       PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT   MOD_RES      760   760       PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT   MOD_RES      767   767       PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT   MOD_RES      817   817       PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT   SITE         516   516       INTERACTION WITH SHC PROTEIN (BY
FT                                SIMILARITY).
FT   SITE         817   817       INTERACTION WITH PLC-GAMMA-1 (BY
FT                                SIMILARITY).
FT   CARBOHYD      87    87       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD      95    95       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD     121   121       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD     178   178       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD     205   205       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD     241   241       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD     254   254       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD     285   285       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD     325   325       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD     329   329       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD     411   411       N-LINKED (GLCNAC...) (POTENTIAL).
FT   VARSPLIC     467   477       FACTIONDEGG -> FYLFHRIFLDG (IN TRUNCATED
FT                                ISOFORM).
FT   VARSPLIC     478   822       MISSING (IN TRUNCATED ISOFORM).
FT   VARIANT      338   338       N -> Y (IN DBSNP:1047856).
FT                                /FTId=VAR_011973.
SQ   SEQUENCE   822 AA;  91698 MW;  2FE8E1B9948E0D15 CRC64;

  Query Match             84.4%;  Score 38;  DB 1;  Length 822;
  Best Local Similarity   80.0%;  Pred. No. 27,
  Matches    4;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 WVRWH 5
           |:|||
Db       4 WIRWH 8


RESULT 6
FD6C_SOYBN
ID   FD6C_SOYBN     STANDARD;      PRT;   434 AA.
AC   P49628;
DT   01-FEB-1996 (Rel. 33, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Omega-6 fatty acid desaturase, chloroplast precursor (EC 1.14.99.-).
OS   Glycine max (Soybean).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.
OX   NCBI_TaxID=3847;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Seed;
RX   MEDLINE=94196008; PubMed=8166133;
RA   Hitz W.D., Carlson T.J., Booth J.R. Jr., Kinney A.J., Shanklin J.,
RA   Yadav N.S.;
RT   "Cloning of a higher-plant plastid omega-6 fatty acid desaturase cDNA
RT   and its expression in a cyanobacterium.";
RL   Plant Physiol. 105:635-641(1994).
CC    !- FUNCTION: CHLOROPLAST OMEGA-6 FATTY ACID DESATURASE INTRODUCED
```

```
CC       THE SECOND DOUBLE BOND IN THE BIOSYNTHESIS OF 16:3 AND 18:3 FATTY
CC       ACIDS, IMPORTANT CONSTITUENTS OF PLANT MEMBRANES. IT IS THOUGHT
CC       TO USE FERREDOXIN AS AN ELECTRON DONOR AND TO ACT ON FATTY ACIDS
CC       ESTERIFIED TO GALACTOLIPIDS, SULFOLIPIDS AND PHOSPHATIDYLGLYCEROL.
CC   -!- PATHWAY: Polyunsaturated fatty acid biosynthesis.
CC   -!- SUBCELLULAR LOCATION: CHLOROPLAST, MEMBRANE-BOUND (PROBABLE).
CC   -!- DOMAIN: THE HISTIDINE BOX DOMAINS MAY CONTAIN THE ACTIVE SITE
CC       AND/OR BE INVOLVED IN METAL ION BINDING.
CC   -!- SIMILARITY: BELONGS TO THE FATTY ACID DESATURASE FAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; L29215; AAA50158.1; -.
DR   InterPro; IPR001225; FA_desaturase.
DR   Pfam; PF00487; FA_desaturase; 1.
DR   ProDom; PD001081; FA_desaturase; 1.
KW   Oxidoreductase; Fatty acid biosynthesis; Chloroplast; Membrane;
KW   Transit peptide.
FT   TRANSIT       1     63       CHLOROPLAST (BY SIMILARITY).
FT   CHAIN        64    424       OMEGA-6 FATTY ACID DESATURASE.
FT   DOMAIN      165    169       HISTIDINE BOX-1.
FT   DOMAIN      201    205       HISTIDINE BOX-2.
FT   DOMAIN      361    365       HISTIDINE BOX-3.
SQ   SEQUENCE   424 AA;  49641 MW;  4F3DF52D4B1A2009 CRC64;

  Query Match             80.0%;  Score 36;  DB 1;  Length 424;
  Best Local Similarity   66.7%;  Pred. No. 30;
  Matches    4;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 WVRWHF 6
           |: |||
Db     251 WLMWHF 256


RESULT 7
BCSB_XANAC
ID   BCSB_XANAC     STANDARD;      PRT;   788 AA.
AC   P58933;
DT   15-JUN-2002 (Rel. 41, Created)
DT   15-JUN-2002 (Rel. 41, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Cyclic di-GMP binding protein precursor (Cellulose synthase regulatory
DE   subunit).
GN   BCSB OR XAC3517.
OS   Xanthomonas axonopodis (pv. citri).
OC   Bacteria; Proteobacteria; Gamma subdivision; Xanthomonas group;
OC   Xanthomonas.
OX   NCBI_TaxID=92829;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=306 / ATCC 13902 / XV 101;
RX   MEDLINE=22022145; PubMed=12024217;
RA   da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,
RA   Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,
RA   Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,
RA   Camarotte G., Cannavan F., Cardozo J., Chambergo F., Ciapina L.P.,
RA   Cicarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorry H.,
RA   Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,
RA   Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,
RA   Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,
RA   Locali E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
RA   Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,
RA   Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.P.,
RA   Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,
RA   Spinola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Terra R.I.D.,
RA   Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,
RA   Setubal J.C., Kitajima J.P.;
```

```
FT   "Comparison of the genomes of two Xanthomonas pathogens with differing
FT   host specificities.",
RL   Nature 417:459-463(2002).
CC   -!- FUNCTION: Binds the cellulose synthase activator, bis-(3',5')-
CC       cyclic diguanylic acid (c-di-GMP) (By similarity).
CC   -!- SUBUNIT: Tightly associated with the cellulose synthase catalytic
CC       subunit (By similarity).
CC   -!- PATHWAY: Bacterial cellulose biosynthesis.
CC   -!- SIMILARITY: BELONGS TO THE ACSB/BCSB FAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AE012000; AAM38360.1; ALT_INIT.
KW   Cellulose biosynthesis; Signal.
FT   SIGNAL        1     34       POTENTIAL.
FT   CHAIN        35    788       CYCLIC DI-GMP BINDING PROTEIN.
SQ   SEQUENCE   788 AA;  86167 MW;  [illegible] CRC64;

  Query Match             80.0%;  Score 36;  DB 1;  Length 788;
  Best Local Similarity   66.7%;  Pred. No. 56;
  Matches    4;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 WVRWHF 6
           |:|||:|
Db     743 WLRWYF 748


RESULT 8
SUV3_DROME
ID   SUV3_DROME     STANDARD;      PRT;  1169 AA.
AC   P20193;
DT   01-FEB-1991 (Rel. 17, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Suppressor of variegation protein 3-7.
GN   SUVAR(3)7.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC   Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC   Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=95223788; PubMed=7708496;
RA   Cleard F., Matsarskaia M., Spierer P.;
RT   "The modifier of position-effect variegation Suvar(3)7 of Drosophila:
RT   there are two alternative transcripts and seven scattered zinc
RT   fingers, each preceded by a tryptophan box.";
RL   Nucleic Acids Res. 23:796-802(1995).
RN   [2]
RP   ERRATUM.
RA   Cleard F., Matsarskaia M., Spierer P.;
RT   Nucleic Acids Res. 23:3804-3804(1995).
RN   [3]
RP   SEQUENCE OF 255-1169 FROM N.A.
RX   MEDLINE=90190836; PubMed=2107402;
RA   Reuter G., Giarre M., Farah J., Gausz J., Spierer A., Spierer P.;
RT   "Dependence of position-effect variegation in Drosophila on dose of a
RT   gene encoding an unusual zinc-finger protein.";
RL   Nature 344:219-223(1990).
CC   -!- FUNCTION: THIS PROTEIN IS A DOSE-LIMITING FACTOR IN POSITION-
CC       EFFECT VARIEGATION, THE INACTIVATION IN SOME CELLS OF A GENE
CC       TRANSLOCATED NEXT TO HETEROCHROMATIN. IT COULD PLAY A ROLE IN
CC       CHROMOSOME CONDENSATION.
CC   -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
```

```
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to licensee@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; X82187; CAA36434.1; -.
DR   PIR; S09151; S09151.
DR   FlyBase; FBgn0003598; Sp[Mus]3.7.
DR   InterPro; IPR000411; BESS_motif.
DR   InterPro; IPR000822; Znf_C2H2.
DR   Pfam; PF02944; BESS; 1.
DR   SMART; SM00355; ZnF_C2H2; 1.
DR   PROSITE; PS00028; ZINC_FINGER_C2H2_1; 1.
KW   Zinc finger; Nuclear protein; DNA binding; Metal-binding; Repeat.
FT   ZN_FING      136    155       C2H2 TYPE.
FT   ZN_FING      238    262       C2H2-TYPE.
FT   ZN_FING      344    368       C2H2-TYPE.
FT   ZN_FING      406    431       C2H2-TYPE.
FT   ZN_FING      526    548       C2H2-TYPE.
FT   ZN_FING      656    680       C2H2 TYPE.
FT   ZN_FING      748    771       C2H2-TYPE.
FT   DOMAIN        86     92       POLY-ASP.
FT   DOMAIN       589    592       POLY-GLU.
FT   DOMAIN       780    788       POLY-ALA.
FT   DOMAIN      1008   1011       POLY-ASN.
FT   DOMAIN      1113   1116       POLY-ASN.
SQ   SEQUENCE   1169 AA;  131113 MW;  ACB9380341648328 CRC64;

  Query Match             80.0%;  Score 36;  DB 1;  Length 1169;
  Best Local Similarity   80.0%;  Pred. No. 82;
  Matches    4;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 WVRWH 5
           |:||:
Db     219 WLRWH 223


RESULT 9
YC07_METJA
ID   YC07_METJA     STANDARD;      PRT;   225 AA.
AC   Q58604;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hypothetical protein MJ1207.
GN   MJ1207.
OS   Methanococcus jannaschii.
OC   Archaea; Euryarchaeota; Methano[illegible];
OC   Methanococcaceae; Methanococcus.
OX   NCBI_TaxID=2190;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=LAL-1 / DSM 2661 / ATCC 43067;
RX   MEDLINE 96387939, PubMed 8688087;
RA   Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA   Sutton G.G., Blake J.A., FitzGerald L.M., Clayton R.A., Gocayne J.D.,
RA   Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA   Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA   Scott J.L., Geoghagen N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA   Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA   Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA   Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.,
RT   "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT   jannaschii.";
RL   Science 273:1058-1073(1996).
CC   -!- SIMILARITY: BELONGS TO THE ACETYLTRANSFERASE FAMILY. STRONG, TO
CC       A.FULGIDUS AF0521.
CC   -----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to licensee@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; U67562, AAB99211.1; -.
DR   TIGR; MJ1207; -.
DR   InterPro; IPR000182, GCN5-acetyltransf.
DR   Pfam; PF00583; Acetyltransf; 1.
KW   Hypothetical protein; Transferase; Acyltransferase; Complete proteome.
SQ   SEQUENCE   225 AA;  25555 MW;  [illegible] CRC64;

  Query Match             77.8%;  Score 35;  DB 1;  Length 226;
  Best Local Similarity   66.7%;  Pred. No. 24;
  Matches    4;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 WVRWHF 6
           | ||:|
Db     109 WARWYF 114


RESULT 10
PLSB_CAEEL
ID   PLSB_CAEEL     STANDARD;      PRT;   716 AA.
AC   Q22949;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   Probable glycerol 3 phosphate acyltransferase, mitochondrial precursor
DE   (EC 2.3.1.15) (GPAT).
GN   F08F3.2.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadoria; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Bristol N2;
RA   Blanchard M., Bradshaw H.;
RL   Submitted (JUL-1998) to the EMBL/GenBank/DDBJ databases.
CC   -!- CATALYTIC ACTIVITY: Acyl CoA + sn glycerol 3 phosphate = CoA + 1-
CC       acyl sn-glycerol 3-phosphate.
CC   -!- PATHWAY: FIRST STEP IN DE NOVO PHOSPHOLIPID BIOSYNTHESIS. IT MAY
CC       ALSO FUNCTION IN THE REGULATION OF MEMBRANE BIOGENESIS.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein. Mitochondrial
CC       (Potential).
CC   -!- SIMILARITY: BELONGS TO THE GPAT / DAPAT FAMILY.
CC   -----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to licensee@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; U64847; AAB04876.1; -.
DR   WormPep; F08F3.2; CE09258.
DR   InterPro; IPR002123; Acyltransferase.
DR   Pfam; PF01553; Acyltransferase; 1.
KW   Phospholipid biosynthesis; Transferase; Acyltransferase;
KW   Transmembrane; Mitochondrion; Transit peptide.
FT   TRANSIT       1      ?       MITOCHONDRION (POTENTIAL).
FT   CHAIN         ?    716       PROBABLE GLYCEROL-3-PHOSPHATE
FT                                ACYLTRANSFERASE.
FT   TRANSMEM    409    425       POTENTIAL.
SQ   SEQUENCE   718 AA;  82071 MW;  E0A56A4A56F0148D CRC64;

  Query Match             77.8%;  Score 35;  DB 1;  Length 718;
  Best Local Similarity   66.7%;  Pred. No. 75;
  Matches    4;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 WVRWHF 6
```

```
          |  |||
Db    178 WCNWHF 183


RESULT 11
RS12_METJA
ID   RS12_METJA      STANDARD;      PRT;   148 AA.
AC   P54062;
DT   01-OCT-1996 (Rel. 34, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   30S ribosomal protein S12P.
GN   RPS12P OR MJ1046
OS   Methanococcus jannaschii.
OC   Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC   Methanocaldococcaceae; Methanocaldococcus.
OX   NCBI_TaxID=2190;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 43067 / DSM 2661 / JAL-1 / JCM 10045 / NBRC 100440;
RX   MEDLINE=96337099; PubMed=8688087;
RA   Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA   Sutton G.G., Blake J.A., FitzGerald L.M., Clayton R.A., Gocayne J.D.,
RA   Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA   Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA   Scott J.L., Geoghagen N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA   Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA   Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA   Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.,
RT   "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT   jannaschii.";
RL   Science 273:1058-1073(1996).
CC   -!- FUNCTION: PROTEIN S12 IS INVOLVED IN THE TRANSLATION INITIATION
CC       STEP (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE S12P FAMILY OF RIBOSOMAL PROTEINS.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; U67547; AAB99050.1; -.
DR   TIGR; MJ1046; -.
DR   InterPro; IPR000529; Ribosomal_S12
DR   Pfam; PF00164; Ribosomal_S12; 1
DR   ProDom; PD000576; Ribosomal_S12; 1
DR   TIGRFAMs; TIGR00982; S23_S12_E_A; 1.
DR   PROSITE; PS00055; RIBOSOMAL_S12; 1.
KW   Ribosomal protein; Complete proteome.
SQ   SEQUENCE   148 AA;  16809 MW;  B2B1744BBDBJCFD CRC64;

  Query Match             75.6%;  Score 34;  DB 1;  Length 148;
  Best Local Similarity   80.0%;  Pred. No. 23;
  Matches    4;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 WVRWH 5
           | |||
Db      22 WCRWH 26


RESULT 12
CBL_ECOLI
ID   CBL_ECOLI       STANDARD;      PRT;   316 AA.
AC   Q47083; P76353;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Transcriptional regulator cbl.
GN   CBL OR B1987.
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
OX   NCBI_TaxID=562;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12;
RX   MEDLINE=96105196; PubMed=8529872;
RA   Iwanicka-Nowicka R., Hryniewicz M.M.;
RT   "A new gene, cbl, encoding a member of the LysR family of
RT   transcriptional regulators belongs to Escherichia coli cys regulon.";
RL   Gene 166:11-17(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12 / MG1655;
RX   MEDLINE=97426617; PubMed=9278503;
RA   Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA   Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA   Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA   Mau B., Shao Y.;
RT   "The complete genome sequence of Escherichia coli K-12.";
RL   Science 277:1453-1474(1997).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12;
RX   MEDLINE=97261158; PubMed=9037040;
RA   Itoh T., Aiba H., Baba T., Fujita K., Hayashi K., Inada T.,
RA   Isono K., Kasai H., Kimura S., Kitakawa M., Kitagawa M.,
RA   Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,
RA   Nakade S., Nakamura Y., Nashimoto H., Nishio Y., Oshima T.,
RA   Saito N., Sampei G., Seki Y., Sivasundaram S., Tagami H.,
RA   Takeda J., Takemoto K., Wada C., Yamamoto Y., Horiuchi T.;
RT   "A 460 kb DNA sequence of the Escherichia coli K-12 genome
RT   corresponding to the 40.1-50.0 min region on the linkage map.";
RL   DNA Res. 3:379-392(1996).
CC   -!- FUNCTION: MAY BE AN ACTIVATORY REGULATORY PROTEIN WITHIN THE CYS
CC       REGULON.
CC   -!- SIMILARITY: BELONGS TO THE LYSR FAMILY OF TRANSCRIPTIONAL
CC       REGULATORS.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; L31629; AAA99029.1; -.
DR   EMBL; AE000290; AAC75049.1; -.
DR   EMBL; D90837; BAA15805.1; -.
DR   HSSP; P45600; 1AL3.
DR   EcoGene; EG14264; cbl.
DR   InterPro; IPR000847; HTH_LysR.
DR   InterPro; IPR005119; LysR_subst
DR   Pfam; PF00126; HTH_1; 1.
DR   Pfam; PF03466; LysR_substrate; 1.
DR   PRINTS; PR00039; HTHLYSR.
DR   PROSITE; PS00044; HTH_LYSR_FAMILY; 1.
KW   Transcription regulation; DNA-binding; Complete proteome.
FT   DNA_BIND     19     39       H-T-H MOTIF (POTENTIAL).
FT   CONFLICT    143    143       A  > R (IN REF. 1).
SQ   SEQUENCE   316 AA;  35856 MW;  70CB594765010670 CRC64;

  Query Match             75.6%;  Score 34;  DB 1;  Length 316;
  Best Local Similarity   80.0%;  Pred. No. 49;
  Matches    4;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 WVRWH 5
           | |||
Db     163 WFRWH 167


RESULT 13
```

```
CBL_KLEAE
ID   CBL_KLEAE      STANDARD;      PRT;   316 AA.
AC   Q08598;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   Transcriptional regulator cbl.
GN   CBL.
OS   Klebsiella aerogenes
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Klebsiella.
OX   NCBI_TaxID=28451;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN W70 / KC1043;
RX   MEDLINE 93203957; PubMed 8458853;
RA   Schwacha A., Bender R.A.;
RT   "The nac (hut put regulator) gene of Klebsiella
RT   aerogenes.";
RL   J. Bacteriol. 175:2107-2115(1993).
CC   -!- FUNCTION: MAY BE AN ACCESSORY REGULATORY PROTEIN WITHIN THE CYS
CC       REGULON.
CC   -!- SIMILARITY: BELONGS TO THE LYSR FAMILY OF TRANSCRIPTIONAL
CC       REGULATORS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaborat
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a licensing agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L01114; AAA19174.1; -.
DR   PIR; C47099; C47099.
DR   HSSP; P45600; 1AL3.
DR   InterPro; IPR000847; HTH_LysR.
DR   InterPro; IPR005119; LysR_subst.
DR   Pfam; PF00126; HTH_1; 1.
DR   Pfam; PF03466; LysR_substrate; 1.
DR   PRINTS; PR00039; HTHLYSR.
DR   PROSITE; PS00044; HTH_LYSR_FAMILY; 1.
KW   Transcription regulation; DNA-binding.
FT   DNA_BIND      19     38       H-T-H MOTIF (POTENTIAL).
SQ   SEQUENCE   316 AA;  35754 MW;  [illegible]

  Query Match           75.6%;  Score 34;  DB 1;  Length 316;
  Best Local Similarity   66.7%;  Pred. No. 49;
  Matches    4;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 WVRWH 5
           | |||
Db     163 WPRWH 167


RESULT 14
MTC1_CAUCR
ID   MTC1_CAUCR     STANDARD;      PRT;   358 AA.
AC   Q45971;
DT   30-MAY-2000 (Rel. 39, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   M.CcrMI [C-5] methylase CcrMI (EC 2.1.1.73) (Adenine specific
DE   methyltransferase CcrMI) (M.CcrMI).
GN   CCRMIM OR CCRM OR CC0378.
OS   Caulobacter crescentus.
OC   Bacteria; Proteobacteria; alpha subdivision; Caulobacter group.
OC   Caulobacter.
OX   NCBI_TaxID=155892;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN CB15N / NA1000;
RX   MEDLINE [illegible], PubMed [illegible],
RA   Zweiger G., Marczynski G., Shapiro L.;
RT   "A Caulobacter DNA methyltransferase that functions only in the
RT   predivisional cell.";
RL   J. Mol. Biol. 235:472-485(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 19089 / CB15;
RX   MEDLINE=21173698; PubMed=11259647;
RA   Nierman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
RA   Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
RA   Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
RA   DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,
RA   Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,
RA   Utterback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,
RA   Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
RT   "Complete genome sequence of Caulobacter crescentus.";
RL   Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
CC   -!- FUNCTION: THIS METHYLASE RECOGNIZES THE 5'-GANTC-3' SEQUENCE
CC       AND CAUSES SPECIFIC METHYLATION ON A-2 ON BOTH STRANDS.
CC       CCRM-MEDIATED METHYLATION HAS IMPORTANT CELLULAR FUNCTIONS.
CC       APPEARS TO CONTRIBUTE TO THE ACCURATE CELL-CYCLE CONTROL OF DNA
CC       REPLICATION AND CELLULAR MORPHOLOGY.
CC   -!- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + DNA adenine = S-
CC       adenosyl-L-homocysteine + DNA 6-methylaminopurine.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaborat
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a licensing agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U01032; AAA18913.1; -.
DR   EMBL; AE005711; AAK22365.1; -.
DR   HSSP; P11409; 1BOO.
DR   REBASE; 2539; M.CcrMI.
DR   TIGR; CC0378; -.
DR   InterPro; IPR001091; CN4_Metransf.
DR   InterPro; IPR002296; Dam_mtPase.
DR   InterPro; IPR002941; N6/N4_Mtase.
DR   InterPro; IPR002052; N6_Mtase.
DR   InterPro; IPR000051; SAM_bind.
DR   Pfam; PF01555; N6_N4_Mtase; 1.
DR   PRINTS; PR00508; S21N4MTFRASE.
DR   PROSITE; PS00092; N6_MTASE; 1.
KW   Transferase; Methyltransferase; DNA replication; Complete proteome.
FT   CONFLICT    242    242       Y -> S (IN REF. 1).
SQ   SEQUENCE   358 AA;  39665 MW;  [illegible];

  Query Match           75.6%;  Score 34;  DB 1;  Length 358;
  Best Local Similarity   66.7%;  Pred. No. 55;
  Matches    4;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 WVRWHF 6
           |  |||
Db     332 WTYWHF 337


RESULT 15
GP45_HUMAN
ID   GP45_HUMAN     STANDARD;      PRT;   372 AA.
AC   Q9Y5Y3;
DT   15-JUN-2002 (Rel. 41, Created)
DT   15-JUN-2002 (Rel. 41, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Probable G protein-coupled receptor GPR45 (PSP24-alpha) (PSP24-1).
GN   GPR45.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
```

```
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99156852; PubMed 10036181;
RA   Marchese A., Sawzdargo M., Nguyen T., Cheng R., Heng H.H., Nowak T.,
RA   Im D.-S., Lynch K.R., George S.R., O'Dowd B.F.;
RT   "Discovery of three novel orphan G-protein-coupled receptors.";
RL   Genomics 56:12-21(1999).
CC   -!- FUNCTION: Orphan receptor. May play a role in brain function.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- TISSUE SPECIFICITY: Expressed in brain; detected in the basal
CC       forebrain, frontal cortex, and caudate, but not in thalamus,
CC       hippocampus, or putamen.
CC   -!- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF118266; AAD21056.1; -.
DR   Genew; HGNC:4503; GPR45.
DR   MIM; 604838; -.
DR   InterPro; IPR000276; GPCR_Rhodpsn.
DR   Pfam; PF00001; 7tm_1; 1.
DR   PRINTS; PR00237; GPCRRHODOPSN.
DR   PROSITE; PS00237; G_PROTEIN_RECEP_F1_1; FALSE_NEG.
DR   PROSITE; PS50262; G_PROTEIN_RECEP_F1_2; 1.
KW   G-protein coupled receptor; Transmembrane; Glycoprotein.
FT   DOMAIN        1     38       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     39     59       1 (POTENTIAL).
FT   DOMAIN       60     75       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     76     96       2 (POTENTIAL).
FT   DOMAIN       97    109       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    110    130       3 (POTENTIAL).
FT   DOMAIN      131    149       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    150    170       4 (POTENTIAL).
FT   DOMAIN      171    198       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    199    219       5 (POTENTIAL).
FT   DOMAIN      220    268       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    269    289       6 (POTENTIAL).
FT   DOMAIN      290    305       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    306    326       7 (POTENTIAL).
FT   DOMAIN      327    372       CYTOPLASMIC (POTENTIAL).
FT   CARBOHYD      4      4       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     17     17       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     20     20       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   372 AA;  42000 MW;  B4B4FC7AEDE20199 CRC64;

  Query Match             75.6%;  Score 34;  DB 1;  Length 372;
  Best Local Similarity   100.0%;  Pred. No. 57;
  Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 VRWHF 6
           |||||
Db      98 VRWHF 102


Search completed: February 20, 2003, 10:48:01
Job time : 8.66667 secs
```

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd

OM protein - protein search, using sw model

Run on: February 21, 2003, 10:45:10 ; Search time 23.3333 Seconds
(without alignments)
52.984 Million cell updates/sec

Title: US-09-912-414-2
Perfect score: 45
Sequence: 1 WVRWHF 6

Scoring table: BLOSUM62DX
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_21:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	41	91.1	219	10	Q94K47	Q94k47 arabidopsis
2	41	91.1	256	10	Q9FIN2	Q9fin2 arabidopsis
3	40	88.9	201	10	Q9c600	Q9c600 arabidopsis
4	40	88.9	202	10	Q9AXY3	Q9axy3 brassica ca
5	40	88.9	213	10	Q9AXY5	Q9axy5 brassica na
6	40	88.9	216	10	Q9AXY2	Q9axy2 brassica na
7	40	88.9	222	10	Q9AXY4	Q9axy4 brassica ol
8	39	86.7	148	16	Q9RDS5	Q9rds5 streptomyce
9	39	86.7	293	16	Q9I065	Q9i065 pseudomonas
10	39	86.7	308	6	Q8WNV6	Q8wnv6 capra hircu
11	39	86.7	370	10	Q9M9G2	Q9m9g2 arabidopsis
12	39	86.7	407	10	Q8RXQ4	Q8rxq4 arabidopsis
13	38	84.4	70	4	Q8WXJ4	Q8wxj4 homo sapien
14	38	84.4	178	6	Q95K65	Q95k65 macaca fasc
15	38	84.4	537	4	Q8WXJ6	Q8wxj6 homo sapien
16	38	84.4	553	4	Q8WXJ5	Q8wxj5 homo sapien
17	38	84.4	838	4	Q8WXJ7	Q8wxj7 homo sapien
18	37	82.2	591	4	Q13041	Q13041 homo sapien
19	36	80.0	185	16	Q9C5Y1	Q9c5y1 pseudomonas
20	36	80.0	223	16	Q98BA4	Q98ba4 rhizobium l
21	36	80.0	249	16	Q5YE04	Q5ye04 mycobacteri
22	36	80.0	262	16	Q9BYG3	Q9byg3 pseudomonas
23	36	80.0	264	16	Q8XZQ5	Q8xzq5 ralstonia s
24	36	80.0	265	7	Q4ZFF5	Q4zff5 pseudomonas
25	36	80.0	265	16	Q8ZB03	Q8zb03 yersinia pe
26	36	80.0	266	16	Q8U649	Q8u649 agrobacteri
27	36	80.0	384	16	Q9RS74	Q9rs74 deinococcus
28	36	80.0	472	16	Q8YMN9	Q8ymn9 anabaena sp
29	36	80.0	754	16	Q8ZQG3	Q8zqg3 salmonella
30	36	80.0	754	16	Q8ZA82	Q8za82 salmonella
31	36	80.0	1169	5	Q9VFX7	Q9vfx7 drosophila
32	36	80.0	1175	5	Q8WT43	Q8wt43 caenorhabdi
33	36	80.0	1178	5	Q8WT44	Q8wt44 caenorhabdi
34	35	77.8	103	11	Q9DCY4	Q9dcy4 mus musculu
35	35	77.8	111	16	Q8ZPL2	Q8zpl2 salmonella
36	35	77.8	111	16	Q8Z6S2	Q8z6s2 salmonella
37	35	77.8	135	11	Q9CFW3	Q9cfw3 mus musculu
38	35	77.8	147	11	Q61427	Q61427 mus musculu
39	35	77.8	190	4	Q96N05	Q96n05 homo sapien
40	35	77.8	458	16	Q9PHQ5	Q9phq5 campylobact
41	35	77.8	494	17	Q974C6	Q974c6 sulfolobus
42	35	77.8	508	17	Q97WR4	Q97wr4 sulfolobus
43	35	77.8	522	5	Q9BHW0	Q9bhw0 leishmania
44	35	77.8	621	5	Q95PS1	Q95ps1 leishmania
45	34	75.6	179	16	Q9I177	Q9i177 pseudomonas

ALIGNMENTS

RESULT 1
Q94K47
ID Q94K47 PRELIMINARY; PRT; 219 AA.
AC Q94K47;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Putative 6-phosphogluconolactonase.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Yamada K., Liu S.X., Sakano H., Pham P.K., Banh J., Chung M.K.,
RA Goldsmith A.D., Lee J.M., Dinh H.L., Tang C.T., Tirtiri M., Y. G.,
RA Bowser L., Carninci P., Chen H., Cheuk R., Hayashizaki Y., Ishida J.,
RA Jones T., Kamiya A., Karlin-Neumann G., Kawai J., Kim C., Koesema E.,
RA Lam B., Lin J., Meyers M.C., Miranda M., Narusaka M., Nguyen M.,
RA Palm C.J., Sakurai T., Satou M., Seki M., Shinn P., Southwick A.,
RA Shinozaki K., Davis R.W., Ecker J.R., Theologis A.;
RT "Full Length cDNA of gene K16H17.11/AT5g24400 (GI:10177901).";
RL Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
CC -!- SIMILARITY: CONTAINS 1 WD REPEAT (TRP ASP DOMAIN).
DR EMBL; AF370305; AAK44120.1; -.
DR TIGRFAMs; TIGR01198; pgl; 1.
KW Repeat; WD repeat.
SQ SEQUENCE 219 AA; 24542 MW; 1994A723D2434B8A CRC64;

Query Match 91.1%; Score 41; DB 10; Length 219;
Best Local Similarity 83.3%; Pred. No. 23;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 WVRWHF 6
| ||||
Db 72 WARWHF 77

```
RESULT 2
Q9FIN2
ID   Q9FIN2      PRELIMINARY;      PRT;   256 AA.
AC   Q9FIN2;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   6 phosphogluconolactonase-like protein.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota, Viridiplantae, Streptophyta, Embryophyta, Tracheophyta,
OC   Spermatophyta, Magnoliophyta, eudicotyledons, core eudicots, Rosidae,
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=COLUMBIA;
RX   MEDLINE=99196233; PubMed=10048486;
RA   Asamizu E., Sato S., Kaneko T., Nakamura Y., Kotani H., Miyajima N.,
RA   Tabata S.;
RT   "Structural analysis of Arabidopsis thaliana chromosome 5. VIII.
RT   Sequence features of the regions of 1,081,959 bp covered by seventeen
RT   physically assigned P1 and TAC clones.";
RL   DNA Res. 5:379-391(1998).
DR   EMBL; AB016884; BAB11023.1; -.
DR   InterPro; IPR000487; Glucosamine_iso.
DR   Pfam; PF01182; Glucosamine_iso; 1.
DR   TIGRFAMs; TIGR01198; pgl; 1.
SQ   SEQUENCE   256 AA;  28034 MW;  21E7046ECFDC72AF CRC64;

  Query Match             91.1%;  Score 41;  DB 10;  Length 256;
  Best Local Similarity   83.3%;  Pred. No. 21;
  Matches    5;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 WVPWHF 6
             | :|||
Db      72 WARWHF 77


RESULT 3
Q9C600
ID   Q9C600      PRELIMINARY;      PRT;   201 AA.
AC   Q9C600;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   Omega 6 reductase (Fragment).
GN   FAD6.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota, Viridiplantae, Streptophyta, Embryophyta, Tracheophyta,
OC   Spermatophyta, Magnoliophyta, eudicotyledons, core eudicots, Rosidae,
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. COLUMBIA;
RA   Fourmann M., Froger N., Brunel D.;
RT   "Amplified consensus gene markers. Tools designing for a genetic map
RT   of Arabidopsis-known function genes in Brassica.";
RL   Submitted (JAN 2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF229332; AAK19064.1; -.
DR   InterPro; IPR001225; FA_desaturase.
DR   Pfam; PF00487; FA_desaturase; 1.
DR   ProDom; PD001081; FA_desaturase; 1.
FT   NON_TER       1       1
FT   NON_TER     201     201
SQ   SEQUENCE   201 AA;  23094 MW;  [illegible] CRC64;

  Query Match             88.9%;  Score 40;  DB 10;  Length 201;
  Best Local Similarity   83.3%;  Pred. No. 30;
  Matches    5;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 WVPWHF 6
             | :|||
Db     156 WVNWHF 161


RESULT 4
Q9AXY3
ID   Q9AXY3      PRELIMINARY;      PRT;   202 AA.
AC   Q9AXY3;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   Omega 6 reductase (Fragment).
GN   FAD6-BR-1.
OS   Brassica campestris (Field mustard).
OC   Eukaryota, Viridiplantae, Streptophyta, Embryophyta, Tracheophyta;
OC   Spermatophyta, Magnoliophyta, eudicotyledons, core eudicots, Rosidae;
OC   eurosids II, Brassicales, Brassicaceae, Brassica.
OX   NCBI_TaxID=3711;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. P909;
RA   Fourmann M., Froger N., Brunel D.;
RT   "Amplified consensus gene markers. Tools designing for a genetic map
RT   of Arabidopsis-known-function genes in Brassica.";
RL   Submitted (JAN 2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF229330; AAK00662.1; -.
DR   InterPro; IPR001225; FA_desaturase.
DR   Pfam; PF00487; FA_desaturase; 1.
DR   ProDom; PD001081; FA_desaturase; 1.
FT   NON_TER       1       1
FT   NON_TER     202     202
SQ   SEQUENCE   202 AA;  23163 MW;  A9CB4C3AAB1BC00M CRC64;

  Query Match             88.9%;  Score 40;  DB 10;  Length 202;
  Best Local Similarity   83.3%;  Pred. No. 30;
  Matches    5;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 WVRWHF 6
             || ||'
Db     157 WVNWHF 162


RESULT 5
Q9AXY5
ID   Q9AXY5      PRELIMINARY;      PRT;   213 AA.
AC   Q9AXY5;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   Omega 6 reductase (Fragment).
GN   FAD6-BN-2.
OS   Brassica napus (Rape).
OC   Eukaryota, Viridiplantae, Streptophyta, Embryophyta, Tracheophyta;
OC   Spermatophyta, Magnoliophyta, eudicotyledons, core eudicots, Rosidae;
OC   eurosids II, Brassicales, Brassicaceae, Brassica.
OX   NCBI_TaxID=3708;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. STELLAR;
RA   Fourmann M., Froger N., Brunel D.;
RT   "Amplified consensus gene markers. Tools designing for a genetic map
RT   of Arabidopsis-known-function genes in Brassica.";
RL   Submitted (JAN 2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF229328; AAK00660.1; -.
DR   InterPro; IPR001225; FA_desaturase.
DR   Pfam; PF00487; FA_desaturase; 1.
DR   ProDom; PD001081; FA_desaturase; 2.
FT   NON_TER       1       1
FT   NON_TER     213     213
SQ   SEQUENCE   213 AA;  24471 MW;  [illegible] CRC64;

  Query Match             88.9%;  Score 40;  DB 10;  Length 213;
  Best Local Similarity   83.3%;  Pred. No. 32;
  Matches    5;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;
```

```
Qy       1 WVRWHF 6
           || |||
Db     154 WVNWHF 159


RESULT 6
Q9AXY2
ID   Q9AXY2      PRELIMINARY;      PRT;   216 AA.
AC   Q9AXY2;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Omega 6 reductase (Fragment).
GN   FAD6 BN 1.
OS   Brassica napus (Rape).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta, Magnoliophyta, eudicotyledons, core eudicots, Rosidae,
OC   eurosids II; Brassicales; Brassicaceae; Brassica.
OX   NCBI_TaxID=3708;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. STELLAR;
RA   Fourmann M., Froger N., Brunel D.;
RT   "Amplified consensus gene markers: Tools designing for a genetic map
RT   of Arabidopsis-known-function genes in Brassica.";
RL   Submitted (JAN 2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF229391; AAF00663.1; -.
DR   InterPro; IPR001225; FA_desaturase.
DR   Pfam; PF00487; FA_desaturase; 1.
DR   ProDom; PD001081; FA_desaturase; 2.
FT   NON_TER       1       1
FT   NON_TER     216     216
SQ   SEQUENCE   216 AA;  25110 MW;  FFFCFFDBRIB0E3B4 CRC64;

  Query Match             88.9%;  Score 40;  DB 10;  Length 216;
  Best Local Similarity   83.3%;  Pred. No. 32;
  Matches    5;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 WVRWHF 6
           || |||
Db     157 WVNWHF 162


RESULT 7
Q9AXY4
ID   Q9AXY4      PRELIMINARY;      PRT;   222 AA.
AC   Q9AXY4;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Omega 6 reductase (Fragment).
GN   FAD6-BO-1.
OS   Brassica oleracea (Cauliflower).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta, Magnoliophyta, eudicotyledons, core eudicots, Rosidae,
OC   eurosids II, Brassicales, Brassicaceae, Brassica.
OX   NCBI_TaxID=3712;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. RAPID CYCLING;
RA   Fourmann M., Froger N., Brunel D.,
RT   "Amplified consensus gene markers: Tools designing for a genetic map
RT   of Arabidopsis-known-function genes in Brassica.";
RL   Submitted (JAN 2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF229389; AAF00661.1; -.
DR   InterPro; IPR001225; FA_desaturase.
DR   Pfam; PF00487; FA_desaturase; 1.
DR   ProDom; PD001081; FA_desaturase; 2.
FT   NON_TER       1       1
FT   NON_TER     222     222
SQ   SEQUENCE   222 AA,  [illegible] MW,  A3B1F8A4A4B7D4E [illegible] CRC64;

  Query Match             88.9%;  Score 40;  DB 10;  Length 222;
  Best Local Similarity   83.3%;  Pred. No. 33;
  Matches    5;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 WVRWHF 6
           || |||
Db     163 WVNWHF 168


RESULT 8
Q9RDS5
ID   Q9RDS5      PRELIMINARY;      PRT;   148 AA.
AC   Q9RDS5;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Putative membrane protein.
GN   SCO2242 OR SC1G2.04C.
OS   Streptomyces coelicolor.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=1902;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2) / M145;
RA   Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
RA   Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
RA   Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA   Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA   Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neil S.,
RA   Rabbinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,
RA   Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA   Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
RA   Hopwood D.A.,
RT   "Complete genome sequence of the model actinomycete Streptomyces
RT   coelicolor A3(2).";
RL   Nature 417:141-147(2002).
DR   EMBL; AL136500; CAB66176.1; -.
SQ   SEQUENCE   148 AA;  16191 MW;  [illegible] CRC64;

  Query Match             86.7%;  Score 39;  DB 16;  Length 148;
  Best Local Similarity   66.7%;  Pred. No. 72;
  Matches    4;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 WVRWHF 6
           |:|||:
Db     125 WLRWHY 130


RESULT 9
Q9I065
ID   Q9I065      PRELIMINARY;      PRT;   293 AA.
AC   Q9I065;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Hypothetical protein PA2778.
GN   PA2778.
OS   Pseudomonas aeruginosa.
OC   Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC   Pseudomonas
OX   NCBI_TaxID=287;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 15692 / PAO1;
RX   MEDLINE=20437337; PubMed=10984043;
RA   Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warrener P.,
RA   Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA   Garber R.L., Goltry L., Tolentino E., Westbrock-Wadman S., Yuan Y.,
RA   Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA   Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA   Reizer J., Saier M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT   "Complete genome sequence of Pseudomonas aeruginosa PAO1, an
```

```
DT   [illegible] pathogen.",
RL   Nature 410 [illegible]
DP   EMBL; AE006470; AAK06166.1; -.
DR   InterPro; IPR005804; Peptidase_C39
DR   InterPro; IPR001440; TPR
DR   Pfam; PF03412; Peptidase_C39; 1.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   [illegible] AA;  [illegible] MW;  [illegible] CRC64;

  Query Match           96.7%;  Score 29;  DB 16;  Length 203;
  Best Local Similarity   93.2%;  Pred. No. 63;
  Matches   5;  Conservative   0;  Mismatches   1;  Indels   0;  Gaps   0;

Qy       1 WVPWHF 6
           | ||||
Db     101 WPRWHF 106


RESULT 10
Q8WVY8
ID   Q8WVY8      PRELIMINARY;      PRT;   308 AA.
AC   Q8WVY8;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Putative pheromone receptor gVIP1.
GN   GVIP1.
OS   Capra hircus (Goat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC   Bovidae; Caprinae; Capra
OX   NCBI_TaxID=9925;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SHIBA;
RA   Wakabayashi Y., Mori Y., Ishikawa M., Yazaki F., Hagio Yamagishi F.,
RT   "Goat gVIP putative pheromone receptor is expressed in the [illegible]
RT   olfactory organs.";
RL   Submitted (OCT-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB064662; BAB93522.1; -.
DR   InterPro; IPR000276; GPCR_Rhodpsn.
DR   InterPro; IPR000126; VIP.
DR   Pfam; PF00002; VIP; 1.
DR   PROSITE; PS50262; G_PROTEIN_RECEP_F1_2; 1.
KW   Receptor.
SQ   SEQUENCE   308 AA;  35379 MW;  [illegible] CRC64;

  Query Match           96.7%;  Score 29;  DB 6;  Length [illegible];
  Best Local Similarity   100.0%;  Pred. No. 66;
  Matches   5;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy       1 WVPWH 5
           |||||
Db     304 WVPWH 309


RESULT 11
Q9M9G2
ID   Q9M9G2      PRELIMINARY;      PRT;   339 AA.
AC   Q9M9G2;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   [illegible] (TrEMBLrel. 15, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   F14G23.22 protein
GN   F14G23.22
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. COLUMBIA;
RA   Liu G., Yu G., Lee J., Sakano H., Bhavani A., Lenz C., Toriumi M.,
RA   Chin C., Chiou J., Choi E., Gonzales A., Howng E., Kim T., Li J.,
RA   Lin A., Pham P., Vaysberg M., Alrafi H., Buehler E., Chan Q., Conn L.,
RA   Conway A.B., Hansen N., Johnson-Hopson C., Khan S., Kim C., Lam B.,
RA   Nguyen M., Palm C., Shinn P., Tambunga G., Davis R., Ecker J.,
RA   Federspiel N., Theologis A.,
RT   "The sequence of BAC F14G23 from Arabidopsis thaliana chromosome 1.";
RL   Submitted (NOV 1999) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. COLUMBIA;
RA   Theologis A.,
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: CONTAINS 7 WD REPEATS (TRP-ASP DOMAINS).
DR   EMBL; AC012654; AAF42237.1; -.
DR   InterPro; IPR001680; WD40.
DR   Pfam; PF00400; WD40; 7.
DR   PRINTS; PR00320; GPROTEINBRPT.
DR   SMART; SM00320; WD40; 7.
DR   PROSITE; PS50082; WD_REPEATS_1; UNKNOWN_1.
DR   PROSITE; PS50294; WD_REPEATS_2; 7.
DR   PROSITE; PS50294; WD_REPEATS_REGION; 1.
KW   Repeat; WD repeat.
SQ   SEQUENCE   [illegible] AA;  [illegible] MW;  [illegible] CRC64;

  Query Match           96.7%;  Score 29;  DB 10;  Length 370;
  Best Local Similarity   100.0%;  Pred. No. 79;
  Matches   5;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy       1 WVPWH 5
           |||||
Db     151 WVPWH 155


RESULT 12
Q9FXQ4
ID   Q9FXQ4      PRELIMINARY;      PRT;   407 AA.
AC   Q9FXQ4;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Hypothetical 43.9 kDa protein
GN   AT1G71840.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Yamada K., Lin S.X., Sakano H., Pham P.K., Banh J., Chung M.K.,
RA   Goldsmith A.D., Lee J.M., Quach H.L., Toriumi M., Yu G., Bowser L.,
RA   Carninci P., Chen H., Chouk P., Hayashizaki Y., Ishida J., Jones T.,
RA   Kamiya A., Karlin-Neumann G., Kawai J., Kim C., Lam P., Lin J.,
RA   Miranda M., Narusaka M., Nguyen M., Palm C.J., Sakurai T., Satou M.,
RA   Seki M., Shinn P., Southwick A., Shinozaki K., Davis R.W., Ecker J.R.,
RA   Theologis A.;
RT   "Arabidopsis Full Length cDNA Clones.",
RL   Submitted (FEB 2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY093792; AAL26602.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   [illegible] AA;  [illegible] MW;  [illegible] CRC64;

  Query Match           96.7%;  Score 29;  DB 10;  Length 407;
  Best Local Similarity   100.0%;  Pred. No. 87;
  Matches   5;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy       1 WVPWH 5
           |||||
Db     160 WVPWH 164


RESULT 13
```

```
Q8WXJ4
ID   Q8WXJ4      PRELIMINARY;      PRT;    70 AA.
AC   Q8WXJ4;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Neurotrophin receptor tyrosine kinase type 2 (Fragment).
GN   NTRK2.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21656983; PubMed=11798182;
RA   Stoilov P., Castren E., Stamm S.;
RT   "Analysis of the Human TrkB Gene Genomic Organization Reveals Novel
RT   TrkB Isoforms, Unusual Gene Length, and Splicing Mechanism.";
RL   Biochem. Biophys. Res. Commun. 290:1054-1065(2002).
DR   EMBL; AF410902; AAL67968.1; -.
DR   InterPro; IPR000372; LRR_Nterm.
DR   Pfam; PF01462; LRRNT; 1.
DR   SMART; SM00013; LRRNT; 1.
KW   Receptor; Kinase.
FT   NON_TER      70      70
SQ   SEQUENCE   70 AA;  7847 MW;  [illegible] CRC64;

  Query Match             84.4%;  Score 38;  DB 4;  Length 70;
  Best Local Similarity   80.0%;  Pred. No. 22;
  Matches    4;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 WVPWH 5
           |:|||
Db       4 WIRWH 8


RESULT 14
Q95K65
ID   Q95K65      PRELIMINARY;      PRT;   178 AA.
AC   Q95K65;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Hypothetical 20.3 kDa protein.
OS   Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
OC   Cercopithecinae; Macaca.
OX   NCBI_TaxID=9541;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=MEDULLA OBLONGATA;
RA   Osada N., Hida M., Kusuda J., Tanuma R., Iseki K., Hirai M., Terao K.,
RA   Suzuki Y., Sugano S., Hashimoto K.;
RT   "Isolation of full-length cDNA clones from macaque brain cDNA
RT   libraries.";
RL   Submitted (JUL-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB066522; BAB62201.1; -.
DR   InterPro; IPR000372; LRR_Nterm.
DR   Pfam; PF01462; LRRNT; 1.
KW   Hypothetical protein.
SQ   SEQUENCE   178 AA;  20242 MW;  A7B5CA4355F5D58F CRC64;

  Query Match             84.4%;  Score 38;  DB 6;  Length 178;
  Best Local Similarity   80.0%;  Pred. No. 55;
  Matches    4;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 WVPWH 5
           |:|||
Db       4 WIRWH 8


RESULT 15
Q8WXJ6
ID   Q8WXJ6      PRELIMINARY;      PRT;   537 AA.
AC   Q8WXJ6;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Neurotrophin receptor tyrosine kinase type 2 truncated isoform.
GN   NTRK2.
OS   Homo sapiens (Human)
OC   Eukaryota, Metazoa, Chordata, Craniata, Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE 21656983; PubMed 11798182;
RA   Stoilov P., Castren E., Stamm S.;
RT   "Analysis of the Human TrkB Gene Genomic Organization Reveals Novel
RT   TrkB Isoforms, Unusual Gene Length, and Splicing Mechanism.";
RL   Biochem. Biophys. Res. Commun. 290:1054-1065(2002).
DP   EMBL; AF410900; AAL67966.1; -.
DP   InterPro, IPR003599, Ig
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR001611; LRR.
DR   InterPro; IPR000483; LRR_Cterm.
DR   InterPro; IPR000372; LRR_Nterm.
DR   Pfam; PF00047; ig; 1.
DP   Pfam; PF00560; LRR; 1.
DP   Pfam, PF01463, LRRCT, 1.
DP   Pfam; PF01462; LRRNT; 1.
DR   SMART; SM00409; IG; 1.
DR   SMART, SM00408, IGc2, 1.
DP   SMART; SM00082; LRRCT; 1.
DR   SMART; SM00013; LRRNT; 1.
KW   Receptor; Kinase.
SQ   SEQUENCE   537 AA;  59166 MW;  5A8FA252A3871CC1 CRC64;

  Query Match             84.4%;  Score 38;  DB 4;  Length 537;
  Best Local Similarity   80.0%;  Pred. No. 1.6e+02;
  Matches    4;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 WVPWH 5
           |:|||
Db       4 WIRWH 8


Search completed: February 20, 2003, 10:49:18
Job time : 25.3333 secs
```

GenCore version 5.1.3

OM protein - protein search, using sw model

Run on: February 20, 2003, 11:24:02 ; Search time 12 Seconds
(without alignments)
48.067 Million cell updates/sec

Title: US-09-912-414-9
Perfect score: 6
Sequence: 1 WXXWXF 6

Scoring table: OLIGOSX
Gapop 60.0 , Gapext 60.0

Searched: 283224 seqs, 96134422 residues

Word size : 0

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : PIR_73:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	6	100.0	21	2	PH1731	Ig heavy chain V r
2	6	100.0	49	2	I48681	185 kDa glycophosp
3	6	100.0	59	2	E82844	hypothetical prote
4	6	100.0	61	2	H91196	hypothetical prote
5	6	100.0	65	2	S23164	light-harvesting p
6	6	100.0	71	2	F84168	hypothetical prote
7	6	100.0	73	2	AH2788	hypothetical prote
8	6	100.0	76	2	T14724	hypothetical prote
9	6	100.0	78	2	S09449	hypothetical prote
10	6	100.0	79	2	A64750	hypothetical prote
11	6	100.0	83	2	A97568	hypothetical prote
12	6	100.0	86	2	T42185	conserved hypothet
13	6	100.0	87	2	H86753	prophage pi2 prote
14	6	100.0	92	2	S37509	Ig kappa chain V r
15	6	100.0	92	2	JQ1627	hypothetical 10.7K
16	6	100.0	93	2	GB7098	hypothetical prote
17	6	100.0	97	2	E84147	hypothetical prote
18	6	100.0	99	2	C84022	hypothetical prote
19	6	100.0	101	2	S73519	probable lipoprote
20	6	100.0	103	2	F82266	conserved hypothet
21	6	100.0	105	2	JE0384	NADH2 dehydrogenas
22	6	100.0	106	2	I64171	hypothetical prote
23	6	100.0	106	2	C82715	conserved hypothet
24	6	100.0	106	2	G72012	transport permease
25	6	100.0	108	2	AG1164	hypothetical prote
26	6	100.0	108	2	S28241	NADH2 dehydrogenas
27	6	100.0	108	2	AG2505	hypothetical prote
28	6	100.0	109	2	F83057	hypothetical prote
29	6	100.0	109	2	AI2029	hypothetical prote
30	6	100.0	110	2	AD0127	conserved hypothet
31	6	100.0	110	2	AG1523	hypothetical prote
32	6	100.0	116	2	P39966	Ig heavy chain pre
33	6	100.0	116	2	G82398	conserved hypothet
34	6	100.0	117	2	A95381	hypothetical prote
35	6	100.0	118	2	S24527	Ig heavy chain V r
36	6	100.0	119	2	S24522	Ig heavy chain V r
37	6	100.0	119	2	S24492	Ig heavy chain V r
38	6	100.0	119	2	S24512	Ig heavy chain V r
39	6	100.0	119	2	S24493	Ig heavy chain V r
40	6	100.0	119	2	S24491	Ig heavy chain V r
41	6	100.0	119	2	S24525	Ig heavy chain V r
42	6	100.0	119	2	S24520	Ig heavy chain V r
43	6	100.0	119	2	S24519	Ig heavy chain V r
44	6	100.0	119	2	S24511	Ig heavy chain V r
45	6	100.0	119	2	S24494	Ig heavy chain V r

ALIGNMENTS

RESULT 1
PH1731
Ig heavy chain V region (clone GCC-11) - mouse (fragment)
C;Species: Mus musculus (house mouse)
C;Date: 24-Feb-1994 #sequence_revision 24-Feb-1994 #text_change 17-Mar-1999
C;Accession: PH1731
R;McHeyzer-Williams, M.G., McLean, M.J., Lalor, P.A., Nossal, G.J.V.
J. Exp. Med. 178, 295-307, 1993
A;Title: Antigen-driven B cell differentiation in vivo.
A;Reference number: PH1675; MUID:93301607; PMID:8315385
A;Accession: PH1731
A;Molecule type: mRNA
A;Residues: 1-21 <MCH>
A;Experimental source: B cell
A;Note: the authors translated the codon ACA for residue 13 as Ala
C;Superfamily: immunoglobulin V region; immunoglobulin homology
C;Keywords: heterotetramer; immunoglobulin

Query Match 100.0%; Score 6; DB 2; Length 21;
Best Local Similarity 50.0%; Pred. No. 7.1e+02;
Matches 3; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 WXXWXF 6
|::|:|
Db 16 WTSWSF 21

RESULT 2
I48681
185 kDa glycophosphoprotein - mouse (fragment)
C;Species: Mus musculus (house mouse)
C;Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 21-Jun-2002
C;Accession: I48681
R;White, M.R.; Hung, M.C.
Oncogene 7, 677-683, 1992
A;Title: Cloning and characterization of the mouse neu promoter.
A;Reference number: I48681; MUID:92228496; PMID:1348855
A;Accession: I48681
A;Status: preliminary; translated from GP/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-49 <RES>
A;Cross-references: EMBL X61796, NID g36184, PIDN CAA43677.1, PID g36187
C;Genetics:
A;Gene: neu
C;Keywords: ATP; phosphoprotein

Query Match 100.0%; Score 6; DB 2; Length 49;
Best Local Similarity 50.0%; Pred. No. 1.4e+03;
Matches 3; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 WXXWXF 6
|::|:|

```
Db      6 WCRWGF 11


RESULT 3
E82844
hypothetical protein XF0126 [imported] - Xylella fastidiosa (strain 9a5c)
C;Species: Xylella fastidiosa
C;Date: 18-Aug-[illegible]  #sequence_revision [illegible] Aug-[illegible] #text_change [illegible] Aug-[illegible]
C;Accession: E82844
R;anonymous, The Xylella fastidiosa Consortium of the Organization [illegible]
Nature 406, 151-157, 2000
A;Title: The genome sequence of the plant pathogen Xylella fastidiosa.
A;Reference number: A82811; MUID:[illegible]; PMID:[illegible]
A;Note: for a complete list of authors see reference number A82828 [illegible]
A;Accession: E82844
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-59 <SIM>
A;Cr[illegible]; GB:AE[illegible], NP_[illegible], NID:g[illegible], PIDN:AAF[illegible], [illegible]
A;Experimental source: strain 9a5c
R;Simpson, A.J.G., Reinach, F.C., Arruda, P., Abreu, F.A., Acencio, M., Alvarenga, R., A
Bitches, M.R.C., Bucho, M.R.P., Camargo, A.A., Camargo, L.E.A., Carraro, D.M., Carrer, H
ao Neto, E., Dockena, C., El Dorry, H., Faldrioni, A.F., Ferreira, A.J.S.
submitted to GenBank, June 2000
A;Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm
J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laigr
chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E
A;Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;
, F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak
A;Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir
M., Tsuhako, M.H., Valadão, I., Van Sluys, M.A., Vencio, R.Z.N., Almeida, D., Vettore, A.L.,
A;Reference number: A69328
A;Contents: annotation
C;Genetics:
A;Gene: XF0126

  Query Match             100.0%;  Score 6;  DB 2;  Length 59;
  Best Local Similarity   [illegible]%;  Pred. No. [illegible];
  Matches    3;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 WXXWXF 6
           |::|:|
Db      54 WRCWLF 59


RESULT 4
H91196
hypothetical protein PC4544 [imported] - Escherichia coli (strain O157:H7, substrain RI
C;Species: Escherichia coli
C;Date: 19-Jul-2001 #sequence_revision 19-Jul-2001 #text_change 19-Jul-2001
C;Accession: H91196
R;Hayashi, T., Makino, K., Ohnishi, M., Kurokawa, K., Ishii, K., Yokoyama, K., Han, C.-
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A;Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genc
A;Reference number: A99629; MUID:[illegible]; PMID:[illegible]
A;Accession: H91196
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-61 <HAY>
A;Cross-references: GB:BA000007; NID:g[illegible]; PIDN:BAB[illegible]; GSPDB:GN[illegible]
A;Experimental source: strain O157:H7, substrain RIMD [illegible]
C;Genetics:
A;Gene: ECs4544

  Query Match             100.0%;  Score 6;  DB 2;  Length 61;
  Best Local Similarity   [illegible]%;  Pred. No. 1.6e+03;
  Matches    3;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 WXXWXF 6
           |::|:|
Db       3 WDRWRF 8


RESULT 5
S23164
light-harvesting protein alpha chain - Prosthecochloris halochloris
N;Alternate names: antenna pigment protein alpha chain
C;Species: Prosthecochloris halochloris
C;Date: 16-Mar-1[illegible] #sequence_revision 16-Mar-1[illegible] #text_change 22-Aug-[illegible]
C;Accession: S23164
R;Wagner-Huber, R., Brunisholz, R.A., Fiesig, T., Frank, G., Suter, F., Zuber, H.
Eur. J. Biochem. 208, 321-325, 1992
A;Title: The primary structure of the antenna polypeptide of Prosthecochloris halochlo
A;Reference number: S23164; MUID:[illegible]; PMID:[illegible]
A;Accession: S23164
A;Molecule type: protein
A;Residues: 1-65 <WAG>
C;Superfamily: light harvesting protein alpha chain
C;Keywords: antenna complex; bacteriochlorophyll; chromophore; light harvesting pol
F,1/Modified site: N-trimylmethionine #status experimental

  Query Match             [illegible]%;  Score 6;  DB 2;  Length 65;
  Best Local Similarity   [illegible]%;  Pred. No. [illegible];
  Matches    3;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 WXXWXF 6
           |::|:|
Db       2 WKLWKF 7


RESULT 6
F84169
hypothetical protein VNG0079H [imported] - Halobacterium sp. NRC-1
C;Species: Halobacterium sp. NRC-1
C;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C;Accession: F84168
R;Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Lasky, S.
; Leithauser, B.; Keller, K.; Cruz, R.; Danson, M.C.; Hough, D.W.; Maddocks, D.G.; Jablon
 Jung, K.H.; Alam, M.; Freitas, T.
Proc. Natl. Acad. Sci. U.S.A. [illegible], [illegible]
A;Authors: Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Lasky, S.
A;Title: Genome sequence of Halobacterium species NRC-1.
A;Reference number: A84168; MUID:[illegible]; PMID:[illegible]
A;Accession: F84168
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-71 <STO>
A;Cross-references: GB:AE004437; NID:g[illegible]; PIDN:AAG[illegible]; GSPDB:GN[illegible]
C;Genetics:
A;Gene: VNG0079H

  Query Match             [illegible]%;  Score 6;  DB 2;  Length 71;
  Best Local Similarity   [illegible]%;  Pred. No. 1.8e+03;
  Matches    3;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 WXXWXF 6
           |::|:|
Db      49 WWLWGF 54


RESULT 7
AH2788
hypothetical protein Atu[illegible] [imported] - Agrobacterium tumefaciens (strain C58, Dupont
C;Species: Agrobacterium tumefaciens
C;Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 11-Jan-2002
C;Accession: AH2788
R;Wood, D.W., Setubal, J.C., Kaul, R., Monks, D., Chen, I., Wood, G.E., Chan, Y., Wag
uirage, G., Gillet, W., Grant, C., Kutyavin, T., Levy, R., Li, M., McClelle
, Karp, P., Romero, P., Zhang, S.
Science 294, 2317-2323, 2001
A;Authors: Yoo, H., Tao, Y., Biddle, P., Jung, M., Krespan, W., Perry, M., Gordon-Kamm, B
ster, E.W.
A;Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A;Reference number: AB2577; PMID:11743193
```

```
A;Accession: AH2788
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-73 <KUR>
A;Cross-references: GB:AE009698; PIDN:AAK42726.1; PID:g17741165; GCPDB:GN00186
A;Experimental source: strain CO8 (Dupont)
C;Genetics:
A;Gene: Atu1727
A;Map position: circular chromosome

  Query Match             100.0%;  Score 6;  DB 2;  Length 73;
  Best Local Similarity    50.0%;  Pred. No. 1.9e+03;
  Matches    3;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 WXXWXF 6
           |::|:|
Db      13 WNLWIF 18


RESULT 8
T14724
hypothetical protein Y1091 - Yersinia pestis plasmid pMT1
C;Species: Yersinia pestis
C;Date: 30-Sep-1999 #sequence_revision 30-Sep-1999 #text_change 24-Sep-1999
C;Accession: T14724; T15006
R;Hu, P.; Elliott, J.; Mcready, P.; Skowronski, E.; Garnes, J.; Kobayashi, A.; Brolin,
submitted to the EMBL Data Library, March 1998
A;Description: Structural organization of virulence determinants in three Yersinia pesti
A;Reference number: Z18168
A;Accession: T14724
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-76 <HUP>
A;Cross-references: EMBL:AF053947; NID:g2996296; PID:g2996363; PIDN:AAC13243.1
R;Lindler, L.E.; Plano, G.V.; Burland, V.; Mayhew, G.F.; Blattner, F.R.
Infect. Immun. 66, 5731-5742, 1998
A;Title: Complete DNA sequence and detailed analysis of the Yersinia pestis KIM5 plasmid
A;Reference number: Z18268; MUID:99043898; PMID:9826348
A;Accession: T15006
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-76 <LIN>
A;Cross-references: EMBL:AF074611; NID:g2891034; PID:g3881083; PIDN:AAC82749.1
C;Genetics:
A;Gene: Y1091
A;Genome: plasmid pMT1

  Query Match             100.0%;  Score 6;  DB 2;  Length 76;
  Best Local Similarity    50.0%;  Pred. No. 1.9e+03;
  Matches    3;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 WXXWXF 6
           |::|:|
Db      50 WRTWNF 55


RESULT 9
S08449
hypothetical protein 9 - spiroplasma virus 1
C;Species: spiroplasma virus 1, SpV1
C;Date: 30-Jun-1991 #sequence_revision 30-Jun-1991 #text_change 07-Dec-1999
C;Accession: S08449
R;Renaudin, J.; Aullo, P.; Vignault, J.C.; Bove, J.M.
Nucleic Acids Res. 18, 1293, 1990
A;Title: Complete nucleotide sequence of the genome of Spiroplasma citri virus SpV1-R8A2
A;Reference number: S08447; MUID:90206799; PMID:2320423
A;Accession: S08449
A;Status: translation not shown
A;Molecule type: DNA
A;Residues: 1-78 <REN>
A;Cross-references: EMBL:X51314; NID:g61933; PIDN:CAA35727.1; PID:g308838
C;Genetics:
A;Genetic code: SGC3

  Query Match             100.0%;  Score 6;  DB 2;  Length 78;
  Best Local Similarity    50.0%;  Pred. No. 2e+03;
  Matches    3;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 WXXWXF 6
           |::|:|
Db      42 WLIWLF 47


RESULT 10
A64750
hypothetical protein b0249 - Escherichia coli (strain K-12)
C;Species: Escherichia coli
C;Date: 12-Sep-1997 #sequence_revision 12-Sep-1997 #text_change 01-Mar-2002
C;Accession: A64750
R;Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Col
.A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A;Title: The complete genome sequence of Escherichia coli K-12.
A;Reference number: A64720; MUID:97426617; PMID:9278503
A;Accession: A64750
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-79 <BLAT>
A;Cross-references: GP:AE000[illegible]; GP:U[illegible]; NID:g[illegible]; PIDN:AAC73452.1; PID:g1786443;
A;Experimental source: strain K-12, substrain MG1655

  Query Match             100.0%;  Score 6;  DB 2;  Length 79;
  Best Local Similarity    50.0%;  Pred. No. 2e+03;
  Matches    3;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 WXXWXF 6
           |::|:|
Db      50 WRAWNF 55


RESULT 11
A97568
hypothetical protein AGR_C_3173 [imported] - Agrobacterium tumefaciens (strain C58, [illegible]
C;Species: Agrobacterium tumefaciens
C;Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 11-Jan-2002
C;Accession: A97568
R;Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman,
A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.;
Science 294, 2323-2328, 2001
A;Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum
A;Reference number: A97359; PMID:11743194
A;Accession: A97568
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-83 <KUR>
A;Cross-references: GB:AE007869; PIDN:AAK87408.1; PID:g15157825; GCPDB:GN00169
C;Genetics:
A;Gene: AGR_C_3173
A;Map position: circular chromosome

  Query Match             100.0%;  Score 6;  DB 2;  Length 83;
  Best Local Similarity    50.0%;  Pred. No. 2.1e+03;
  Matches    3;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 WXXWXF 6
           |::|:|
Db      23 WNLWIF 28


RESULT 12
T42185
conserved hypothetical protein L7085 - Escherichia coli plasmid pO157
C;Species: Escherichia coli
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C;Accession: T42185
R;Burland, V.; Shao, Y.; Perna, N.T.; Plunkett, G.; Sofia, H.J.; Blattner, F.R.
```

Nucleic Acids Res. 26, 4196-4204, 1998
A;Title: The complete DNA sequence and analysis of the large virulence plasmid of B.bur
A;Reference number: Z22068; MUID:98391744; PMID:9722640
A;Accession: T42185
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-96 <BUR>
A;Cross-references: EMBL:AF074613; PIDN:AAC70183.1
A;Experimental source: strain 7[illegible], serotype B
C;Genetics:
A;Genome: plasmid pQ107
A;Note: L7165

Query Match 100.0%; Score 6; DB 2; Length 96;
Best Local Similarity 50.0%; Pred. No. 2.1e+03;
Matches 3; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 WXXWXF 6
|::|:|
Db 62 WRTWNF 67

RESULT 13
H86753
prophage p12 protein 25 [imported] - Lactococcus lactis subsp. lactis (strain IL1403)
C;Species: Lactococcus lactis subsp. lactis
C;Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 03-Aug-2001
C;Accession: H86753
R;Bolotin, A.; Wincker, P.; Mauger, S.; Jaillon, O.; Malarme, K.; Weissenbach, J.; Ehrl
Genome Res. 11, 731-753, 2001
A;Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis ss
A;Reference number: A86625; MUID:21335186; PMID:11337471
A;Accession: H86753
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-87 <STO>
A;Cross-references: GB:AE001176; PID:g12723983; PIDN:AAK05130.1; GSPDB:GN00146
A;Experimental source: strain IL1403
C;Genetics:
A;Gene: pil25

Query Match 100.0%; Score 6; DB 2; Length 87;
Best Local Similarity 50.0%; Pred. No. 2.2e+03;
Matches 3; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 WXXWXF 6
|::|:|
Db 17 WSAWAF 22

RESULT 14
S37509
Ig kappa chain V region (V kappa 2) - human (fragment)
C;Species: Homo sapiens (man)
C;Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 23-Jul-1999
C;Accession: S37509
R;Klein, U.; Kueppers, R.; Rajewsky, K.
submitted to the EMBL Data Library, September 1993
A;Description: Human IgM(+)IgD(+) cells, the major B cell subset in the peripheral blood
A;Reference number: S37501
A;Accession: S37509
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-92 <KLE>
A;Cross-references: EMBL:Z26003; NID:g406659; PIDN:CAA81356.1; PID:g406660
C;Superfamily: immunoglobulin V region; immunoglobulin homology
C;Keywords: heterotetramer; immunoglobulin

Query Match 100.0%; Score 6; DB 2; Length 92;
Best Local Similarity 50.0%; Pred. No. 2.3e+03;
Matches 3; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 WXXWXF 6
|::|:|
Db 77 WFTWTF 82

RESULT 15
JQ1627
hypothetical 22.7K protein - turkey rhinotracheitis virus
C;Species: turkey rhinotracheitis virus
C;Date: 17-Apr-1998 #sequence_revision 17-Apr-1998 #text_change 08-Oct-1999
C;Accession: JQ1627
R;Ling, R.; Easton, A.J.; Pringle, C.R.
J. Gen. Virol. 72, 1709-1715, 1991
A;Title: Sequence analysis of the 22K, SH and G genes of turkey rhinotracheitis virus an
A;Reference number: A1405; MUID:91333185; PMID:1653697
A;Accession: JQ1627
A;Molecule type: mRNA
A;Residues: 1-92 <LIN>
A;Cross-references: GB:X54193; NID:g581606; PIDN:CAA38118.1; PID:g2981606

Query Match 100.0%; Score 6; DB 2; Length 92;
Best Local Similarity 50.0%; Pred. No. 2.3e+03;
Matches 3; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 WXXWXF 6
|::|:|
Db 65 WAYWFF 70

Search completed: February 21, 2003, 11:29:40
Job time : 13 secs

OM protein - protein search, using sw model

Run on: February 20, 2003, 11:20:17 ; Search time 11.2333 Seconds
(without alignments)
21.958 Million cell updates/sec

Title: US-09-912-414-9
Perfect score: 6
Sequence: 1 WXXWXF 6

Scoring table: OLIGODX
Gapop 60.0 , Gapext 60.0

Searched: 112892 seqs, 41476328 residues

Word size : 0

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	6	100.0	65	1	LHA2_ECTHL	P80103 ectothiorho
2	6	100.0	78	1	VG9_SPV1R	P15900 spiroplasma
3	6	100.0	79	1	YKFF_ECOLI	P75677 escherichia
4	6	100.0	84	1	ACHA_CROPS	P54248 crocidura r
5	6	100.0	84	1	ACHA_FELCA	P54350 felis silve
6	6	100.0	90	1	YFJI_ECOLI	P58096 escherichia
7	6	100.0	101	1	YG50_MYCPN	P75147 mycoplasma
8	6	100.0	105	1	NIGM_HUMAN	P05178 homo sapien
9	6	100.0	106	1	YQCC_HAEIN	Q57152 haemophilus
10	6	100.0	108	1	NIGM_BOVIN	Q02374 bos taurus
11	6	100.0	116	1	HV05_CARAU	P19181 carassius a
12	6	100.0	128	1	MERT_STAAU	P08656 staphylococ
13	6	100.0	129	1	FV4R_HUMAN	P04062 homo sapien
14	6	100.0	129	1	YHW2_YEAST	P38857 saccharomyc
15	6	100.0	138	1	YA91_MYCPN	P75602 mycoplasma
16	6	100.0	139	1	YE13_MYCPN	Q9exd6 mycoplasma
17	6	100.0	144	1	HV26_MOUSE	P01795 mus musculu
18	6	100.0	145	1	PL32_ARPPE	Q9yf92 aeropyrum p
19	6	100.0	146	1	YG5F_YEAST	P53311 saccharomyc
20	6	100.0	147	1	Y565_METJA	Q57985 methanococ
21	6	100.0	155	1	HOPD_ECOLI	O68932 escherichia
22	6	100.0	157	1	YE63_MYCPN	P75320 mycoplasma
23	6	100.0	164	1	LSPA_ECOLI	P00804 escherichia
24	6	100.0	165	1	DSBB_VIBAL	Q56578 vibrio algi
25	6	100.0	165	1	LSPA_ENTAE	P13514 enterobacte
26	6	100.0	165	1	YRH1_RHOSH	Q53229 rhodobacter
27	6	100.0	166	1	Y18P_SSV1	P20209 sulfolobus
28	6	100.0	170	1	LSPA_PSEFL	P17942 pseudomonas
29	6	100.0	171	1	Y428_MYCGE	P47667 mycoplasma
30	6	100.0	172	1	YB03_MYCPN	P75566 mycoplasma
31	6	100.0	176	1	VPI_BPP2	P26701 bacteriopha
32	6	100.0	177	1	TN18_HUMAN	Q9ung2 homo sapien
33	6	100.0	182	1	VG37_BPMU	Q9t1v8 bacteriopha
34	6	100.0	196	1	YJR8_YEAST	P46989 saccharomyc
35	6	100.0	200	1	YE21_MYCTH	P71669 mycobacteri
36	6	100.0	207	1	YNJF_ECOLI	P76226 escherichia
37	6	100.0	208	1	YHHN_ECOLI	P37616 escherichia
38	6	100.0	210	1	SSH5_YEAST	Q03446 saccharomyc
39	6	100.0	213	1	IF4F_XENLA	P48597 xenopus lae
40	6	100.0	215	1	IF4E_APLCA	O77210 aplysia cal
41	6	100.0	217	1	IF4F_HUMAN	P06730 homo sapien
42	6	100.0	217	1	IF4E_MOUSE	P20415 mus musculu
43	6	100.0	217	1	IF4E_RABIT	P09328 oryctolagus
44	6	100.0	219	1	ENGR_METJA	Q57768 methanococ
45	6	100.0	220	1	PP28_BPSP1	P03048 bacteriopha

ALIGNMENTS

RESULT 1
LHA2_ECTHL
ID LHA2_ECTHL STANDARD; PRT; 65 AA.
AC P80103;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 01-FEB-1994 (Rel. 28, Last annotation update)
DE Light-harvesting protein B800/830/1020, alpha-2 chain (EHS-alpha 2)
DE (Antenna pigment protein, alpha-2 chain).
OS Ectothiorhodospira halochloris.
OC Bacteria, Proteobacteria, gamma subdivision, Ectothiorhodospiraceae;
OC Halorhodospira
OX NCBI_TaxID=1052;
RN [1]
RP SEQUENCE.
RC STRAIN=DSM 1059;
RX MEDLINE=92249336; PubMed=1577099;
RA Wagner-Huber R., Brunisholz R.A., Bissig I., Frank G., Suter F.,
RA Zuber H.;
RT "The primary structure of the antenna polypeptides of
RT Ectothiorhodospira halochloris and Ectothiorhodospira halophila. Four
RT core-type antenna polypeptides in E. halochloris and E. halophila.";
RL Eur. J. Biochem. 205:917-925(1992).
CC -!- FUNCTION: ANTENNA COMPLEXES ARE LIGHT-HARVESTING SYSTEMS, WHICH
CC TRANSFER THE EXCITATION ENERGY TO THE REACTION CENTERS.
CC -!- SUBUNIT: THE CORE COMPLEX IS FORMED BY DIFFERENT ALPHA AND BETA
CC CHAINS, BINDING BACTERIOCHLOROPHYLL MOLECULES, AND ARRANGED
CC MOST PROBABLY IN TETRAMERIC STRUCTURES DISPOSED AROUND THE
CC REACTION CENTER. THE NON-PIGMENTED GAMMA CHAINS MAY CONSTITUTE
CC ADDITIONAL COMPONENTS.
CC -!- SUBCELLULAR LOCATION: Type II membrane protein. Inner membrane.
DR PIR; S23164; S23164.
DR InterPro; IPR002361; AntennaA.
DR InterPro; IPR000066; Antenna_a/b.
DR Pfam; PF00556; LHC; 1.
DR PROSITE; PS00968; ANTENNA_COMP_ALPHA; 1.
KW Antenna complex; Light-harvesting polypeptide; Transmembrane;
KW Magnesium; Bacteriochlorophyll; Inner membrane.
FT DOMAIN 1 13 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 14 34 POTENTIAL.
FT DOMAIN 35 65 PERIPLASMIC (POTENTIAL).
FT METAL 29 29 AXIAL LIGAND TO THE BACTERIOCHLOROPHYLL
FT MAGNESIUM (POTENTIAL).
SQ SEQUENCE 65 AA; 7688 MW; 55A4C306/48E3D9A CRC64;

Query Match 100.0%; Score 6; DB 1; Length 65;
Best Local Similarity 50.0%; Pred. No. 7.8e+02;
Matches 3; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 WXXWXF 6
           |..|.|
Db       2 WKLWKF 7
```

RESULT 2
VG9_SPV1R

```
ID   VG3_SPV1B      STANDARD;      PRT;    79 AA.
AC   P15900;
DT   01-APR-1990 (Rel. 14, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   01-FEB-1996 (Rel. 33, Last annotation update)
DE   Gene 9 protein
GN   9.
OS   Spiroplasma virus SpV1 P9A2 B
OC   Viruses; ssDNA viruses; Inoviridae; Plectrovirus.
OX   NCBI_TaxID=10854;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90206799; PubMed=2320423;
RA   Renaudin J., Aullo P., Vignault J.C., Bove J.M.;
RT   "Complete nucleotide sequence of the genome of Spiroplasma citri
RT   virus SpV1 P9A2 B.";
RL   Nucleic Acids Res. 18:1293-1293(1990).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X01344; CAA25727.1; -.
SQ   SEQUENCE   79 AA;  9761 MW;  EF3FE1DBEB7C75B CRC64;

  Query Match             100.0%;  Score 6;  DB 1;  Length 79;
  Best Local Similarity    60.0%;  Pred. No. 9e+02;
  Matches    3;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 WXXWXF 6
             |::|:|
Db      42 WLIWLF 47


RESULT 3
YKFF_ECOLI
ID   YKFF_ECOLI     STANDARD;      PRT;    79 AA
AC   P76677; Q9R2D8;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hypothetical protein ykfF
GN   YKFF OR B0249.
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriales;
OC   Escherichia.
OX   NCBI_TaxID=562;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12 / MG1655;
RX   MEDLINE=97420617; PubMed=9278503;
RA   Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA   Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA   Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA   Mau B., Shao Y.;
RT   "The complete genome sequence of Escherichia coli K-12.";
RL   Science 277:1453-1474(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12 / W3110;
RA   Tabata S., Mori H., Matsuyama N., Hatasha K., Yano M., Ito T.,
RA   Yamamoto Y., Imokuchi H., Miki T., Harada F., Fukuda R., Ichihara S.,
RA   Mizuno T., Makino K., Nakata A., Yura T., Gamper S., Mizobuch K.,
RT   "Systematic sequencing of the Escherichia coli genome: analysis of the
RT   4.0 - 6.0 min (40,007 - 78,416bp) region.",
RL   Submitted (FEB-1996) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: STRONG, TO E.COLI YPJI.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE000123; AAC73252.1; -.
DR   EMBL; D83036; BAA77918.1; ALT_INIT.
DR   EcoGene; EG14283; ykfF.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   79 AA;  9014 MW;  e14b5/4a7/c1A234A CRC64;

  Query Match             100.0%;  Score 6;  DB 1;  Length 79;
  Best Local Similarity    60.0%;  Pred. No. 9.1e+02;
  Matches    3;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 WXXWXF 6
             |::|:|
Db      50 WRAWNF 55


RESULT 4
ACHA_CROSC
ID   ACHA_CROSC     STANDARD;      PRT;    84 AA.
AC   P54248;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Acetylcholine receptor protein, alpha chain (Fragment).
GN   CHRNA1.
OS   Crocidura russula (White-toothed shrew).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Insectivora; Soricidae; Crocidurinae; Crocidura.
OX   NCBI_TaxID=36802;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Muscle;
RX   MEDLINE=95345091; PubMed=7619817;
RA   Barchan D., Ovadia M., Kochva E., Fuchs S.;
RT   "The binding site of the nicotinic acetylcholine receptor in animal
RT   species resistant to alpha-bungarotoxin.";
RL   Biochemistry 34:9172-9176(1995).
CC   -!- FUNCTION: AFTER BINDING ACETYLCHOLINE, THE ALPHA SUBUNIT UNDERGOES AN
CC       EXTENSIVE CHANGE IN CONFORMATION THAT AFFECTS ALL SUBUNITS AND
CC       LEADS TO OPENING OF AN ION-CONDUCTING CHANNEL ACROSS THE PLASMA
CC       MEMBRANE.
CC   -!- SUBUNIT: PENTAMER OF TWO ALPHA CHAINS, AND ONE EACH OF THE BETA,
CC       DELTA, AND GAMMA (OR EPSILON) CHAINS IN IMMATURE MUSCLE. IN MATURE
CC       MUSCLE, GAMMA CHAINS ARE REPLACED BY EPSILON CHAINS.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein (By similarity).
CC   -!- SIMILARITY: BELONGS TO THE LIGAND-GATED ION CHANNEL FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U17006; AAA89113.1; -.
DR   InterPro; IPR001175; Neur_channel.
DR   Pfam; PF02931; Neur_chan_LBD; 1.
DR   PRINTS; PR00254; NEURALRECEPTR; 1.
KW   Receptor; Post-translational modification; Ionic channel; Glycoprotein;
KW   Transmembrane.
FT   NON_TER       1      1
FT   DISULFID      7     21       BY SIMILARITY.
FT   DISULFID      ?      ?       ASSOCIATED WITH RECEPTOR ACTIVATION (BY
FT                                 SIMILARITY).
FT   CARBOHYD     20     20       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   NON_TER      84     84
```

```
SQ   SEQUENCE   84 AA;  9833 MW;  213072FFC967C86B CRC64;

  Query Match             100.0%;  Score 6;  DB 1;  Length 84;
  Best Local Similarity    50.0%;  Pred. No. 9.5e+02;
  Matches    3;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 WXXWXF 6
           |::|:|
Db      63 WKHWVF 68


RESULT 5
ACHA_FELCA
ID   ACHA_FELCA     STANDARD;      PRT;    84 AA.
AC   P54250;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Acetylcholine receptor protein, alpha chain (Fragment)
GN   CHRNA1.
OS   Felis silvestris catus (Cat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Carnivora; Fissipedia; Felidae; Felis.
OX   NCBI_TaxID=9685;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Muscle;
RX   MEDLINE=95345091; PubMed=7619817;
RA   Barchan D., Ovadia M., Kochva E., Fuchs S.;
RT   "The binding site of the nicotinic acetylcholine receptor in animal
RT   species resistant to alpha-bungarotoxin.";
RL   Biochemistry 34:9172-9176(1995).
CC   -!- FUNCTION: AFTER BINDING ACETYLCHOLINE, THE AChR RESPONDS BY AN
CC       EXTENSIVE CHANGE IN CONFORMATION THAT AFFECTS ALL SUBUNITS AND
CC       LEADS TO OPENING OF AN ION-CONDUCTING CHANNEL ACROSS THE PLASMA
CC       MEMBRANE.
CC   -!- SUBUNIT: PENTAMER OF TWO ALPHA CHAINS, AND ONE EACH OF THE BETA,
CC       DELTA, AND GAMMA (IN IMMATURE MUSCLE) OR EPSILON (IN MATURE
CC       MUSCLE) CHAINS.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein (By similarity).
CC   -!- SIMILARITY: BELONGS TO THE LIGAND-GATED ION CHANNEL FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U17007; AAA89114.1; -.
DR   InterPro; IPR001175; Neur_channel.
DR   Pfam; PF02931; Neur_chan_LBD; 1.
DR   PROSITE; PS00236; NEUROTR_ION_CHANNEL; 1.
KW   Receptor; Postsynaptic membrane; Ionic channel; Glycoprotein;
KW   Transmembrane.
FT   NON_TER       1      1
FT   DISULFID      7     21       BY SIMILARITY.
FT   DISULFID     71     72       ASSOCIATED WITH RECEPTOR ACTIVATION (BY
FT                                SIMILARITY).
FT   CARBOHYD     20     20       N-LINKED (GLCNAC...) (POTENTIAL)
FT   NON_TER      84     84
SQ   SEQUENCE   84 AA;  9833 MW;  5B6F9A108A1FBD7F CRC64;

  Query Match             100.0%;  Score 6;  DB 1;  Length 84;
  Best Local Similarity    50.0%;  Pred. No. 9.5e+02;
  Matches    3;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 WXXWXF 6
           |::|:|
Db      63 WKHWVF 68
```

```
RESULT 6
YPJI_ECOLI
ID   YPJI_ECOLI     STANDARD;      PRT;    90 AA.
AC   P58095;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Hypothetical protein ypjI.
GN   YPJI OR B2641.1.
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
OX   NCBI_TaxID=562;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12 / MG1655;
RX   MEDLINE=97426617; PubMed=9278503;
RA   Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA   Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA   Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA   Mau B., Shao Y.;
RT   "The complete genome sequence of Escherichia coli K-12.";
RL   Science 277:1453-1474(1997).
RN   [2]
RP   IDENTIFICATION, AND CONCEPTUAL TRANSLATION.
RA   Rudd K.E.;
RL   Unpublished observations (JAN-1999).
CC   -!- SIMILARITY: STRONG, TO E.COLI YKFF.
CC   -!- CAUTION: THIS IS A CONCEPTUAL TRANSLATION. A FRAMESHIFT HAD TO BE
CC       CORRECTED IN POSITION 69 AND TWO IN FRAME STOP CODONS HAD TO BE
CC       SUPPRESSED IN POSITIONS 8 AND 11.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE000349; -; NOT_ANNOTATED_CDS.
DR   EcoGene; EG13309; ypjI.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   90 AA;  10070 MW;  E1C3E2FFC2EAB683 CRC64;

  Query Match             100.0%;  Score 6;  DB 1;  Length 90;
  Best Local Similarity    50.0%;  Pred. No. 1e+03;
  Matches    3;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 WXXWXF 6
           |::|:|
Db      61 WRAWNF 66


RESULT 7
YG50_MYCPN
ID   YG50_MYCPN     STANDARD;      PRT;   101 AA.
AC   P75147;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hypothetical lipoprotein MPN650 precursor (E09_orf101).
GN   MPN650 OR MP192.
OS   Mycoplasma pneumoniae.
OC   Bacteria, Firmicutes, Mollicutes; Mycoplasmataceae; Mycoplasma.
OX   NCBI_TaxID=2104;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 29342 / M129;
RX   MEDLINE=97105885; PubMed=8948633;
RA   Himmelreich R., Hilbert H., Plagens H., Pirkl E., Li B.-C.,
RA   Herrmann R.;
RT   "Complete sequence analysis of the genome of the bacterium Mycoplasma
```

```
DT   [illegible] (Rel. [illegible]),
RL   Nucleic Acids Res. 24:4420-4449(1996).
CC   -!- SUBCELLULAR LOCATION: Attached to the membrane by a lipid anchor
CC       (Potential).
CC   -!- SIMILARITY: BELONGS TO THE MG438 / MG440 FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaborat
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AB000820; AAB96848.1; -.
DR   InterPro; IPR001595; Lipoprotein_3.
DR   Pfam; PF00938; Lipoprotein_3; 1.
DR   ProDom; PD003276; Lipoprotein_3; 1.
DR   PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.
KW   Hypothetical protein; Lipoprotein; Membrane; Signal;
KW   Complete proteome.
FT   SIGNAL       1     19       POTENTIAL.
FT   CHAIN       20    101       HYPOTHETICAL LIPOPROTEIN MTH610.
FT   LIPID       20     20       N-ACYL DIGLYCERIDE (CYSTEINE).
SQ   SEQUENCE   101 AA;  11183 MW;  C71F0B9EF197573D CRC64;

  Query Match             100.0%;  Score 6;  DB 1;  Length 101;
  Best Local Similarity   50.0%;  Pred. No. 1.1e+03;
  Matches    3;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 WXXWXF 6
             |:::|:
Db      95 WTNWKF 100


RESULT 8
NIGM_HUMAN
ID   NIGM_HUMAN     STANDARD;      PRT;   105 AA.
AC   O95178;
DT   15-JUL-1999 (Rel. 38, Created)
DT   15-JUL-1999 (Rel. 38, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   NADH ubiquinone oxidoreductase AGGG subunit, mitochondrial precursor
DE   (EC 1.6.5.3) (EC 1.6.99.3) (Complex I AGGG) (CI-AGGG).
GN   NDUFB2.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE 95307151; PubMed=7774578;
RA   Loeffen J.L.C.M., Triepels R.H., van den Heuvel L., Schuelke M.,
RA   Buskens C.A.F., Smeets R.J.P., Trijbels J.M.F., Smeitink J.A.M.,
RT   "cDNA of eight nuclear encoded subunits of NADH ubiquinone
RT   oxidoreductase: human complex I cDNA characterized and analysed.",
RL   Biochem. Biophys. Res. Commun. 209:418-423(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Blood;
RX   MEDLINE 20499260; PubMed=11042152;
RA   Zhang Q.-H., Ye M., Wu X.-Y., Ren S.-X., Zhao M., Zhao C.-J., Fu G.,
RA   Shen Y., Fan H.-Y., Lu G., Zhong M., Xu X.-R., Han Z.-G., Zhang J.-W.,
RA   Tao J., Huang Q.-H., Zhou J., Hu G.-X., Gu J., Chen S.-J., Chen Z.;
RT   "Cloning and functional analysis of cDNAs with open reading frames f.;
RT   300 previously undefined genes expressed in CD34+ hematopoietic
RT   stem/progenitor cells.",
RL   Genome Res. 10:1546-1560(2000).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Lung;
RA   Strausberg R.;
RL   Submitted (SEP-2000) to the EMBL/GenBank/DDBJ databases.

CC   -!- FUNCTION: TRANSFER OF ELECTRONS FROM NADH TO THE RESPIRATORY
CC       CHAIN. THE IMMEDIATE ELECTRON ACCEPTOR FOR THE ENZYME IS BELIEVED
CC       TO BE UBIQUINONE.
CC   -!- CATALYTIC ACTIVITY: NADH + ubiquinone = NAD(+) + ubiquinol.
CC   -!- CATALYTIC ACTIVITY: NADH + a quinone = NAD(+) + a reduced acceptor.
CC   -!- SUBUNIT: COMPLEX I IS COMPOSED OF ABOUT 46 DIFFERENT SUBUNITS.
CC   -!- SUBCELLULAR LOCATION: Mitochondrial inner membrane, matrix side.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF000220; AAD05428.1; -.
DR   EMBL; AF067422; AAD02450.1; -.
DR   EMBL; BC001169; AAH01168.1; -.
DR   HGNC; 7697; NDUFB2.
DR   MIM; 603848; -.
KW   Oxidoreductase; NAD; Ubiquinone; Mitochondrion; Transit peptide.
FT   TRANSIT      1     33       MITOCHONDRION (BY SIMILARITY).
FT   CHAIN       34    105       NADH UBIQUINONE OXIDOREDUCTASE AGGG
FT                               SUBUNIT.
SQ   SEQUENCE   105 AA;  12044 MW;  5A0E0F7A0A6CB6B6F CRC64;

  Query Match             100.0%;  Score 6;  DB 1;  Length 105;
  Best Local Similarity   50.0%;  Pred. No. 1.1e+03;
  Matches    3;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 WXXWXF 6
             |:::|:
Db      68 WILWRF 73


RESULT 9
YQCC_HAEIN
ID   YQCC_HAEIN     STANDARD;      PRT;   106 AA.
AC   Q57152; Q05061;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   15-OCT-2002 (Rel. 41, Last annotation update)
DE   Hypothetical protein HI1436.
GN   HI1436.
OS   Haemophilus influenzae.
OC   Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC   Haemophilus.
OX   NCBI_TaxID=727;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Rd; KW20 / ATCC 51907;
RX   MEDLINE 95350619; PubMed=7542800;
RA   Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA   Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA   McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA   Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA   Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA   Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA   Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA   Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA   Venter J.C.;
RT   "Whole-genome random sequencing and assembly of Haemophilus influenzae
RT   Rd.";
RL   Science 269:496-512(1995).
CC   -!- SIMILARITY: TO THE N-TERMINAL OF E.COLI AMIDOPHOSPHORIBOSYLTRANSFERASE
CC       PRECURSOR (EC). THE C-TERMINAL PART IS COLINEAR WITH YQCB.
CC   -!- SIMILARITY: BELONGS TO B.SUBTILIS YLXJ.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
```

```
CC   modified and this statement is not removed.  Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; U32822; AAC23085.1; -.
DR   TIGR; HI1436; -.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   106 AA;  12277 MW;  [illegible] CRC64;

  Query Match             100.0%;  Score 6;  DB 1;  Length 106;
  Best Local Similarity    50.0%;  Pred. No. 1.1e+03;
  Matches    3;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 WXXWXF 6
           |::|:|
Db      48 WLQWVF 53


RESULT 10
NIGM_BOVIN
ID   NIGM_BOVIN     STANDARD;      PRT;   109 AA.
AC   Q02374;
DT   01-JUL-1993 (Rel. 26, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   NADH-ubiquinone oxidoreductase AGGG subunit, mitochondrial precursor
DE   (EC 1.6.5.3) (EC 1.6.99.3) (Complex I AGGG) (CI AGGG).
GN   NDUFB2.
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC   Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 37-65.
RC   TISSUE=Heart;
RX   MEDLINE=92290317; PubMed=1518044;
RA   Walker J.E., Arizmendi J.M., Dupuis A., Fearnley I.M., Finel M.,
RA   Medd S.M., Pilkington S.J., Runswick M.J., Skehel J.M.;
RT   "Sequences of 20 subunits of NADH:ubiquinone oxidoreductase from
RT   bovine heart mitochondria. Application of a novel strategy for
RT   sequencing proteins using the polymerase chain reaction.";
RL   J. Mol. Biol. 226:1051-1072(1992).
CC   -!- FUNCTION: TRANSFER OF ELECTRONS FROM NADH TO THE RESPIRATORY
CC       CHAIN. THE IMMEDIATE ELECTRON ACCEPTOR FOR THE ENZYME IS BELIEVED
CC       TO BE UBIQUINONE.
CC   -!- CATALYTIC ACTIVITY: NADH + ubiquinone = NAD(+) + ubiquinol.
CC   -!- CATALYTIC ACTIVITY: NADH + acceptor = NAD(+) + reduced acceptor.
CC   -!- SUBUNIT: COMPLEX I IS COMPOSED OF ABOUT 40 DIFFERENT SUBUNITS.
CC   -!- SUBCELLULAR LOCATION: Mitochondrial inner membrane; matrix side.
CC   -----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; X63216; CAA44901.1; -.
DR   PIR; S28241; S28241.
KW   Oxidoreductase; NAD; Ubiquinone; Mitochondrion; Transit peptide.
FT   TRANSIT       1     36       MITOCHONDRION
FT   CHAIN        37    109       NADH UBIQUINONE OXIDOREDUCTASE AGGG
FT                                 SUBUNIT.
SQ   SEQUENCE   109 AA;  12294 MW;  [illegible] CRC64;

  Query Match             100.0%;  Score 6;  DB 1;  Length 109;
  Best Local Similarity    50.0%;  Pred. No. 1.2e+03;
  Matches    3;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 WXXWXF 6
           |::|:|
Db      71 WILWRF 76


RESULT 11
HV05_CARAU
ID   HV05_CARAU     STANDARD;      PRT;   116 AA.
AC   P19181;
DT   01-NOV-1990 (Rel. 16, Created)
DT   01-NOV-1990 (Rel. 16, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   Ig heavy chain V region 5A precursor.
OS   Carassius auratus (Goldfish).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC   Cyprinidae; Carassius.
OX   NCBI_TaxID=7957;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=88144476; PubMed=3125551;
RA   Wilson M.R., Middleton D., Warr G.W.;
RT   "Immunoglobulin heavy chain variable region gene evolution: structure
RT   and family relationships of two genes and a pseudogene in a teleost
RT   fish.";
RL   Proc. Natl. Acad. Sci. U.S.A. 85:1566-1570(1988).
DR   PIR; B28966; B28966.
DR   HSSP; P01772; 2FB4.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR003596; Ig_v.
DR   Pfam; PF00047; ig; 1.
DR   SMART; SM00406; IGv; 1.
KW   Immunoglobulin V region; Signal.
FT   SIGNAL        1     19
FT   CHAIN        20    116       IG HEAVY CHAIN V REGION 5A.
FT   DOMAIN       20     49       FRAMEWORK-1.
FT   DOMAIN       50     54       COMPLEMENTARITY-DETERMINING 1.
FT   DOMAIN       55     68       FRAMEWORK 2.
FT   DOMAIN       69     84       COMPLEMENTARITY-DETERMINING 2.
FT   DOMAIN       85    116       FRAMEWORK-3.
FT   DISULFID     41    114       BY SIMILARITY.
FT   NON_TER     116    116
SQ   SEQUENCE   116 AA;  12909 MW;  [illegible] CRC64;

  Query Match             100.0%;  Score 6;  DB 1;  Length 116;
  Best Local Similarity    50.0%;  Pred. No. 1.3e+03;
  Matches    3;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 WXXWXF 6
           |::|:|
Db       4 WLSWVF 9


RESULT 12
MERT_STAAU
ID   MERT_STAAU     STANDARD;      PRT;   128 AA.
AC   P08656;
DT   01-JAN-1988 (Rel. 06, Created)
DT   01-JAN-1988 (Rel. 06, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   Mercuric transport protein (Mercury ion transport protein).
GN   MERT.
OS   Staphylococcus aureus.
OG   Plasmid pI258.
OC   Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX   NCBI_TaxID=1280;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=87260937; PubMed=3037534;
RA   Laddaga R.A., Chu L., Misra T.K., Silver S.;
RT   "Nucleotide sequence and expression of the mercurial-resistance
RT   operon from Staphylococcus aureus plasmid pI258.";
RL   Proc. Natl. Acad. Sci. U.S.A. 84:5106-5110(1987).
CC   -!- FUNCTION: INVOLVED IN MERCURIC TRANSPORT. PASSES A HG(2+) ION
CC       FROM THE PERIPLASMIC MERP PROTEIN TO THE MERCURIC REDUCTASE
```

```
CC       (MERA).
CC    -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC    ---------------------------------------------------------------------
CC    This SWISS-PROT entry is copyright. It is produced through a collaboration
CC    between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC    the European Bioinformatics Institute. There are no restrictions on its
CC    use by non-profit institutions as long as its content is in no way
CC    modified and this statement is not removed. Usage by and for commercial
CC    entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC    or send an email to license@isb-sib.ch).
CC    ---------------------------------------------------------------------
DR    EMBL; L29436; AAA98244.1; -.
DR    PIR; D29564; D29564.
KW    Transport; Mercuric resistance; Transmembrane; Mercury; Plasmid.
FT    TRANSMEM     28     48       POTENTIAL.
FT    TRANSMEM     71     91       POTENTIAL.
FT    METAL        47     47       HG(2+) (POTENTIAL).
FT    METAL        48     48       HG(2+) (POTENTIAL).
FT    METAL       114    114       HG(2+) (POTENTIAL).
FT    METAL       115    115       HG(2+) (POTENTIAL).
SQ    SEQUENCE   128 AA;  14182 MW;  C9974E2B4A8AFE91 CRC64;

  Query Match             100.0%;  Score 6;  DB 1;  Length 128;
  Best Local Similarity   50.0%;  Pred. No. 1.3e+03;
  Matches    3;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 WXXWXF 6
           |::|:|
Db      28 WGVWAF 33


RESULT 13
KV3H_HUMAN
ID   KV3H_HUMAN     STANDARD;      PRT;   123 AA.
AC   P04207;
DT   20-MAR-1987 (Rel. 04, Created)
DT   01-NOV-1990 (Rel. 16, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   Ig kappa chain V-III region CLL precursor (Rheumatoid factor).
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=86177570; PubMed=3083417;
RA   Jirik F.R., Sirge J., Fong S., Wallmann J.G., Clad J.G., Chen P.P.,
RA   Goldfien R., Carson D.A.;
RT   "Cloning and sequence determination of a human rheumatoid factor
RT   light chain gene.";
RL   Proc. Natl. Acad. Sci. U.S.A. 83:6165-6169(1986).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M12740; AAA36990.1; -.
DR   PIR; A01898; K3HUCL.
DR   HSSP; P80362; 1WTL.
DR   InterPro; IPR003596; Ig_MHC.
DR   InterPro; IPR003599; Ig_V.
DR   Pfam; PF00047; ig; 1.
DR   SMART; SM00406; IGv; 1.
KW   Immunoglobulin V region; Signal.
FT   SIGNAL       1     20
FT   CHAIN       21    129       IG KAPPA CHAIN V-III REGION CLL.
FT   DOMAIN      21     43       FRAMEWORK-1.
FT   DOMAIN      44     54       COMPLEMENTARITY DETERMINING-1.
FT   DOMAIN      55     69       FRAMEWORK-2.
FT   DOMAIN      70     76       COMPLEMENTARITY DETERMINING 2.
FT   DOMAIN      77    104       FRAMEWORK-3.
FT   DOMAIN     105    116       COMPLEMENTARITY DETERMINING-3.
FT   DOMAIN     117    127       JOIN SEGMENT.
FT   DISULFID    43    108       BY SIMILARITY.
FT   NON_TER    129    129
SQ   SEQUENCE   129 AA;  14271 MW;  1C139411DB66C714 CRC64;

  Query Match             100.0%;  Score 6;  DB 1;  Length 129;
  Best Local Similarity   50.0%;  Pred. No. 1.3e+03;
  Matches    3;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 WXXWXF 6
           |::|:|
Db     114 WPPWTF 119


RESULT 14
YHW2_YEAST
ID   YHW2_YEAST     STANDARD;      PRT;   189 AA.
AC   P38857;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hypothetical 14.6 kDa protein in PET114-SLA1 intergenic region.
GN   YHR162W.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288c / AB972;
RA   MEDLINE=94178466; PubMed=8139554;
RA   Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Dover J.,
RA   Du Z., Favello A., Fulton L., Gattung S., Geisel C., Kirsten J.,
RA   Kucaba T., Hillier L., Jier M., Johnston L., Langston Y.,
RA   Latreille P., Louis E.J., Macri C., Mardis E., Menezes S., Mouser L.,
RA   Nhan M., Rifkin L., Riles L., St Peter H., Trevaskis E., Vaughan K.,
RA   Vignati D., Wilcox L., Wohldman P., Waterston R., Wilson R.,
RA   Vaudin M.;
RT   "Complete nucleotide sequence of Saccharomyces cerevisiae chromosome
RT   VIII.";
RL   Science 265:2077-2082(1994).
CC   -!- SIMILARITY: BELONGS TO THE UPF0041 (0-44) FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U00027; AAB68009.1; -.
DR   PIR; S48301, S48302.
DR   SGD; S0001205; YHR162W.
DR   InterPro; IPR002906; UPF0041.
DR   Pfam; PF02590; UPF0041; 1.
KW   Hypothetical protein.
SQ   SEQUENCE   129 AA;  14455 MW;  8C1B83A463DB24A7 CRC64;

  Query Match             100.0%;  Score 6;  DB 1;  Length 129;
  Best Local Similarity   50.0%;  Pred. No. 1.3e+03;
  Matches    3;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 WXXWXF 6
           |::|:|
Db      64 WTRWSF 69


RESULT 15
YA91_MYCPN
```

```
ID   YA91_MYCPN     STANDARD;      PRT;   138 AA.
AC   P75602;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hypothetical protein MPN091 (R02_orf138).
GN   MPN091 OR MP064.
OS   Mycoplasma pneumoniae.
OC   Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX   NCBI_TaxID=2104;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 29342 / M129;
RX   MEDLINE=97105885; PubMed=8948633;
RA   Himmelreich R., Hilbert H., Plagens H., Pirkl E., Li B.-C.,
RA   Herrmann R.;
RT   "Complete sequence analysis of the genome of the bacterium Mycoplasma
RT   pneumoniae.";
RL   Nucleic Acids Res. 24:4420-4449(1996).
CC   -!- SIMILARITY: ALMOST IDENTICAL TO M.PNEUMONIAAE MPN413 AND MPN463.
CC   ---------------------------------------------------------- ----------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AE000008; AAB95711.1; -.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   138 AA;  15763 MW;  F5238D5428F42709 CRC64;

  Query Match             100.0%;  Score 6;  DB 1;  Length 138;
  Best Local Similarity    50.0%;  Pred. No. 1.4e+03;
  Matches    3;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 WXXWXF 6
             |::|:|
Db      36 WSGWSF 41


Search completed: February 20, 2003, 11:24:57
Job time : 12.3333 secs
```

GenCore version 5.1.3

OM protein - protein search, using sw model

Run on: February 20, 2003, 11:23:27 ; Search time 46 Seconds
(without alignments)
26.876 Million cell updates/sec

Title: US-09-912-414-9
Perfect score: 6
Sequence: 1 WXXWXF 6

Scoring table: OLIGODX
Gapop 60.0 , Gapext 60.0

Searched: 671580 seqs, 206047115 residues

Word size : 0

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : SPTREMBL_21:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	6	100.0	28	6	O62821	O62821 bubalus bub
2	6	100.0	31	8	Q94NH2	Q94nh2 litterina s
3	6	100.0	36	15	Q905D7	Q905d7 human immun
4	6	100.0	48	11	Q63365	Q63365 rattus norv
5	6	100.0	49	11	Q08615	Q08615 mus musculu
6	6	100.0	50	4	Q9NPC0	Q9npc0 homo sapien
7	6	100.0	59	16	Q9PH20	Q9ph20 xylella fas
8	6	100.0	61	16	O85619	O85619 escherichia
9	6	100.0	64	4	Q8WVM9	Q8wvm9 homo sapien
10	6	100.0	70	2	Q9EXE4	Q9exe4 lactobacill
11	6	100.0	71	17	Q9HST8	Q9hst8 halobacteri
12	6	100.0	74	2	Q50168	Q50168 mycobacteri
13	6	100.0	76	16	O68793	O68793 yersinia pe
14	6	100.0	77	2	Q9S4W3	Q9s4w3 escherichia
15	6	100.0	79	2	Q9WTF7	Q9wtf7 escherichia
16	6	100.0	79	16	Q9JMR3	Q9jmr3 escherichia
17	6	100.0	80	16	Q93GP6	Q93gp6 salmonella
18	6	100.0	83	16	Q8UEM9	Q8uem9 agrobacteri
19	6	100.0	85	2	Q9F5S9	Q9f5s9 escherichia
20	6	100.0	86	2	Q9ZGR9	Q9zgr9 escherichia
21	6	100.0	86	10	Q8W1M2	Q8w1m2 solanum cha
22	6	100.0	87	16	Q9CGR4	Q9cgr4 lactococcus
23	6	100.0	89	10	Q8W1L9	Q8w1l9 solanum pin
24	6	100.0	90	10	Q8W1L8	Q8w1l8 solanum pin
25	6	100.0	92	10	Q8W1M0	Q8w1m0 solanum bul
26	6	100.0	92	12	Q90090	Q90090 turkey rhin
27	6	100.0	93	10	Q8W1M1	Q8w1m1 solanum bul
28	6	100.0	93	16	Q9CBW2	Q9cbw2 mycobacteri
29	6	100.0	95	4	Q9Y6T4	Q9y6t4 homo sapien
30	6	100.0	96	4	Q96QD1	Q96qd1 homo sapien
31	6	100.0	97	16	Q9K5V6	Q9k5v6 bacillus ha
32	6	100.0	99	10	Q94AN0	Q94an0 arabidopsis
33	6	100.0	99	16	Q9KBM7	Q9kbm7 bacillus ha
34	6	100.0	101	5	Q95RS1	Q95rs1 drosophila
35	6	100.0	103	10	Q9XJ52	Q9xj52 triticum ae
36	6	100.0	103	16	Q9KTL5	Q9ktl5 vibrio chol
37	6	100.0	103	16	Q8XXI0	Q8xxi0 ralstonia s
38	6	100.0	105	11	Q9CPU2	Q9cpu2 mus musculu
39	6	100.0	106	16	Q9CNF4	Q9cnf4 pasteurella
40	6	100.0	106	16	Q9PE50	Q9pe50 xylella fas
41	6	100.0	106	16	Q9K1V7	Q9k1v7 chlamydia p
42	6	100.0	106	16	Q9Z6T9	Q9z6t9 chlamydia p
43	6	100.0	107	2	O06658	O06658 rhizobium l
44	6	100.0	108	16	Q8YKR8	Q8ykr8 anabaena sp
45	6	100.0	108	16	Q8Y823	Q8y823 listeria mo

ALIGNMENTS

RESULT 1
O62821
ID O62821 PRELIMINARY; PRT; 28 AA.
AC O62821;
DT 01-AUG-1998 (TrEMBLrel. 07, Created)
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Lysozyme (Fragment).
OS Bubalus bubalis (Domestic water buffalo).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC Bovidae; Bovinae; Bubalus.
OX NCBI_TaxID=89462;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BLOOD;
RA Das P., Tiwari G., Garg L.C.;
RT "Characterisation of lysozyme encoding gene promoter in Bubalus
RT bubalis.";
RL Submitted (MAR-1998) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AJ225012; CAA12342.1; -.
FT NON_TER 28 28
SQ SEQUENCE 28 AA; 3160 MW; B8CE6F6AD7DDB29D CRC64;

Query Match 100.0%; Score 6; DB 6; Length 28;
Best Local Similarity 50.0%; Pred. No. 1.7e+03;
Matches 3; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 WXXWXF 6
           |..|:|
Db      10 WSVWTF 15
```

RESULT 2
Q94NH2
ID Q94NH2 PRELIMINARY; PRT; 31 AA.
AC Q94NH2;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)

```
DT   01-MAR 2002 (TrEMBLrel. 20, Last annotation update)
DE   Cytochrome oxidase subunit II (Fragment)
OS   Littorina [illegible].
OG   Mitochondrion.
OC   Eukaryota; Metazoa; Mollusca; Gastropoda; Prosobranchia;
OC   Mesogastropoda; Littorinoidea; Littorinidae; Littorina.
OX   NCBI_TaxID=31220;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Moran P., Rolan Alvarez E.;
RL   Submitted (SEP-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF416614; AAL17810.1; -.
DR   EMBL; AF416615; AAL17812.1; -.
DR   InterPro; IPR002429; Cyt_c_ox_2.
DR   Pfam; PF02790; COX2_TM; 1.
KW   Mitochondrion.
FT   NON_TER      31     31
SQ   SEQUENCE   31 AA;  3613 MW;  [illegible] CRC64;

  Query Match             [illegible]%;  Score 6;  DB 1;  Length 31;
  Best Local Similarity   [illegible]%;  Pred. No. [illegible];
  Matches    3;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 WXXWXP 6
           |::|:|
Db       4 WGQWGP 9


RESULT 3
Q9C5D7
ID   Q9C5D7      PRELIMINARY;      PRT;    36 AA.
AC   Q9C5D7;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   Envelope glycoprotein (Fragment).
OS   Human immunodeficiency virus type 1
OC   Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX   NCBI_TaxID=11676;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=[illegible];
RA   Taniguchi Y., Takehisa J., Bikandou B., Mboudjeka I.,
RA   N'Doundi M'Pidi M. Y., M'Pandi M., M'Pele P., Hurida Y., Hayami M.,
RA   Ichimura H., Parra H. J.;
RT   "Genetic Subtypes of HIV Type 1 Found in the gp160env Sequences in
RT   Republic of Congo Brazzaville.";
RL   Submitted (OCT-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF410454; AAL10285.1; -.
FT   NON_TER      36     36
SQ   SEQUENCE   36 AA;  4434 MW;  [illegible] CRC64;

  Query Match             [illegible]%;  Score 6;  DB 1;  Length 36;
  Best Local Similarity   [illegible]%;  Pred. No. [illegible];
  Matches    3;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 WXXWXP 6
           |::|:|
Db      16 WGIWIP 21


RESULT 4
Q63365
ID   Q63365      PRELIMINARY;      PRT;    48 AA.
AC   Q63365;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01 NOV 1996 (TrEMBLrel. 01, Last sequence update)
DT   01-MAR 2002 (TrEMBLrel. 20, Last annotation update)
DE   Transforming oncogene NEU (Fragment).
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91301737, PubMed=2132832;
RA   Suen T.C., Hung M.C.;
RT   "Multiple cis- and trans-acting elements involved in regulation of the
RT   neu gene.";
RL   Mol. Cell. Biol. 10:6306-6315(1990).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91172172, PubMed=1672439;
RA   Yan D.H., Hung M.C.;
RT   "Identification and characterization of a novel enhancer for the rat
RT   neu promoter.";
RL   Mol. Cell. Biol. 11:1875-1882(1991).
DR   EMBL; M61004; AAA41666.1; -.
KW   Oncogene, Transforming protein.
FT   NON_TER      48     48
SQ   SEQUENCE   48 AA;  4[illegible] MW;  [illegible] CRC64;

  Query Match             [illegible]%;  Score 6;  DB 11;  Length 48;
  Best Local Similarity   [illegible]%;  Pred. No. [illegible];
  Matches    3;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 WXXWXP 6
           |::|:|
Db       9 WCRWGP 14


RESULT 5
Q08615
ID   Q08615      PRELIMINARY;      PRT;    49 AA.
AC   Q08615;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV 1996 (TrEMBLrel. 01, Last sequence update)
DT   01 NOV 1999 (TrEMBLrel. 12, Last annotation update)
DE   Avian ERYTHROBLASTOSIS oncogene B 2 (EC 2.7.1.112) (NEU oncogene)
DE   (P185) (Epidermal growth factor receptor related protein) (Fragment).
GN   ERBB2 OR NEU.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92226496, PubMed=1348855;
RA   White M.K.-A., Hung M.-C.;
RT   "Cloning and characterization of the mouse neu promoter.";
RL   Oncogene 7:677-683(1992).
CC   -!- CATALYTIC ACTIVITY: ATP + A PROTEIN TYROSINE = ADP + PROTEIN
CC       TYROSINE PHOSPHATE.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC   -!- SIMILARITY: EQUIVALENT TO HUMAN CERB B-2/HER-2.
CC   -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
DR   EMBL; X68220; CAA48466.1; -.
DR   MGD; MGI:99419; Erbb2.
KW   Transmembrane; Glycoprotein; Duplication; Receptor; Transferase;
KW   Tyrosine protein kinase; ATP binding; Phosphorylation; Oncogene.
FT   NON_TER      49     49
SQ   SEQUENCE   49 AA;  [illegible] MW;  [illegible] CRC64;

  Query Match             100.0%;  Score 6;  DB 11;  Length 49;
  Best Local Similarity   50.0%;  Pred. No. 2.7e-03;
  Matches    3;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 WXXWXP 6
           |::|:|
Db       6 WCRWGP 11


RESULT 6
Q9NRC0
ID   Q9NRC0      PRELIMINARY;      PRT;    50 AA.
AC   Q9NRC0;
```

```
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   FAM4A1 (Fragment).
GN   FAM4A1.
OS   Homo sapiens (Human)
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20349292; PubMed=10889047;
RA   Vincent J.B., Herbrick J.A., Gurling H.M.D., Bolton P.F., Roberts W.,
RA   Scherer S.W.;
RT   "Identification of a novel gene on chromosome 7q31 that is interrupted
RT   by a translocation breakpoint in an autistic individual.";
RL   Am. J. Hum. Genet. 67:510-514(2001).
DR   EMBL; AF234886; AAF85949.1; -.
FT   NON_TER      50      50
SQ   SEQUENCE   50 AA;  5941 MW;  9AFBB51E74DBEC83 CRC64;

  Query Match             100.0%;  Score 6;  DB 4;  Length 50;
  Best Local Similarity   50.0%;  Pred. No. 2.7e+03;
  Matches    3;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 WXXWXF 6
           |::|:|
Db      24 WTVWPF 29


RESULT 7
Q9PH20
ID   Q9PH20      PRELIMINARY;      PRT;    59 AA.
AC   Q9PH20;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Hypothetical protein XF0126.
GN   XF0126.
OS   Xylella fastidiosa.
OC   Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
OC   Xylella.
OX   NCBI_TaxID=2371;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=9A5C;
RX   MEDLINE=20365717; PubMed=10910347;
RA   Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,
RA   Alvarenga R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,
RA   Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,
RA   Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carrer H.,
RA   Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
RA   Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorry H.,
RA   Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,
RA   Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furlan L.R.,
RA   Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
RA   Ho P.L., Hoheisel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,
RA   Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,
RA   Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
RA   Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,
RA   Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,
RA   Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
RA   Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
RA   Nhani A. Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,
RA   de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
RA   Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,
RA   Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
RA   de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,
RA   da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,
RA   da Silveira J.F., Silvestri M.Z., Siqueira W.J., de Souza A.A.,
RA   de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tsuhako M.H.,
RA   Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
RA   Zago M.A., Zatz M., Meidanis J., Setubal J.C.;
RT   "The genome sequence of the plant pathogen Xylella fastidiosa.";
RL   Nature 406:151-159(2000).
DR   EMBL; AE003866; AAF82939.1; -.
DR   InterPro; IPR001230; Prenyl_site.
DR   PROSITE; PS00294; PRENYLATION; UNKNOWN_1.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   59 AA;  7368 MW;  EE[illegible] CRC64;

  Query Match             100.0%;  Score 6;  DB 16;  Length 59;
  Best Local Similarity   50.0%;  Pred. No. 3.1e+03;
  Matches    3;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 WXXWXF 6
           |::|:|
Db      54 WRCWLF 59


RESULT 8
O85619
ID   O85619      PRELIMINARY;      PRT;    61 AA.
AC   O85619;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   L0012 (Hypothetical 5.5 kDa protein).
GN   Z5095 OR ECS4544.
OS   Escherichia coli, and
OS   Escherichia coli O157:H7.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
OX   NCBI_TaxID=562, 83334;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   SPECIES=E.coli; STRAIN=EDL933;
RX   MEDLINE=98229885; PubMed=9673266;
RA   Perna N.T., Mayhew G.F., Posfai G., Elliott S., Donnenberg M.S.,
RA   Kaper J.B., Blattner F.R.;
RT   "Molecular evolution of a pathogenicity island from enterohemorrhagic
RT   escherichia coli O157:H7.";
RL   Infect. Immun. 66:3810-3817(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   SPECIES=E.coli O157:H7; STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX   MEDLINE=21074935; PubMed=11206551;
RA   Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA   Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA   Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA   Grotbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
RA   Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA   Welch R.A., Blattner F.R.;
RT   "Genome sequence of enterchaemorrhagic Escherichia coli O157:H7.";
RL   Nature 409:529-533(2001).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   SPECIES=E coli O157:H7; STRAIN=O157:H7 / RIMD 0509952;
RX   MEDLINE=21156231; PubMed=11258796;
RA   Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA   Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA   Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA   Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT   "Complete genome sequence of enterohemorrhagic Escherichia coli
RT   O157:H7 and genomic comparison with a laboratory strain K-12.";
RL   DNA Res. 8:11-22(2001).
DR   EMBL; AF071034; AAC31491.1; -.
DR   EMBL; AP005594; AAG58810.1; ALT_INIT.
DR   EMBL; AP002566; BAB37967.1; -.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   61 AA;  7286 MW;  5B7F4811F9D58CD3 CRC64;

  Query Match             100.0%;  Score 6;  DB 16;  Length 61;
  Best Local Similarity   50.0%;  Pred. No. 3.2e+03;
  Matches    3;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy      1 WXXWXF 6
          |::|:|
Db      3 WDRWRF 8


RESULT 9
Q8WVM9
ID   Q8WVM9      PRELIMINARY;      PRT;    64 AA.
AC   Q8WVM9;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Hypothetical 7.0 kDa protein (Fragment).
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia, Eutheria, Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BRAIN;
RA   Strausberg R.;
RL   Submitted (DEC-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC017733; AAH17733.1; -.
KW   Hypothetical protein.
FT   NON_TER       1       1
SQ   SEQUENCE   64 AA;  7480 MW;  [illegible] CRC64;

  Query Match             100.0%;  Score 6;  DB 4;  Length 64;
  Best Local Similarity   50.0%;  Pred. No. 3.4e+03;
  Matches    3;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 WXXWXF 6
          |::|:|
Db     16 WDPWLF 21


RESULT 10
Q9EXE4
ID   Q9EXE4      PRELIMINARY;      PRT;    70 AA.
AC   Q9EXE4;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Putative thymidylate synthase (Fragment).
GN   THY.
OS   Lactobacillus sakei.
OC   Bacteria; Firmicutes, Bacilli, [illegible] group, Lactobacillales,
OC   Lactobacillaceae, Lactobacillus.
OX   NCBI_TaxID=1599;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SAK3;
RA   Basso A., Campanile C., Ferrara L., Damiani G.;
RT   "Sequencing and analysis of DNA fragments obtained by random amplified
RT   polymorphic DNA (RAPD) technique from Lactobacillus sakei strains.";
RL   Submitted [illegible] to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AJ404622; CAC20646.1; -.
DR   HSSP; P00469; 1TSY.
DR   InterPro; IPR000398; Thymidylat_synt.
DR   Pfam; PF00303; thymidylat_synt; 1.
DR   ProDom; PD001180; Thymidylat_synt; 1.
FT   NON_TER       1       1
FT   NON_TER      70      70
SQ   SEQUENCE   70 AA;  8483 MW;  [illegible] CRC64;

  Query Match             100.0%;  Score 6;  DB 2;  Length 70;
  Best Local Similarity   50.0%;  Pred. No. 3.6e+03;
  Matches    3;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 WXXWXF 6
          |::|:|
Db     60 WDEWAF 65


RESULT 11
Q9HST8
ID   Q9HST8      PRELIMINARY;      PRT;    71 AA.
AC   Q9HST8;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   Vng0079h.
GN   VNG0079H.
OS   Halobacterium sp. (strain NRC-1).
OC   Archaea; Euryarchaeota; Halobacteria; Halobacteriales;
OC   Halobacteriaceae; Halobacterium.
OX   NCBI_TaxID=64091;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20504493; PubMed=11016950;
RA   Ng W.V., Kennedy S.P., Mahairas G.G., Berquist B., Pan M.,
RA   Shukla H.D., Lasky S.R., Baliga N.S., Thorsson V., Sbrogna J.,
RA   Swartzell S., Weir D., Hall J., Dahl T.A., Welti R., Goo Y.A.,
RA   Leithauser B., Keller K., Cruz R., Danson M.J., Hough D.W.,
RA   Maddocks D.G., Jablonski P.E., Krebs M.P., Angevine C.M., Dale H.,
RA   Isenbarger T.A., Peck R.F., Pohlschroder M., Spudich J.L., Jung K.-H.,
RA   Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,
RA   Ebhardt H., Lowe T.M., Liang P., Riley M., Hood L., DasSarma S.;
RT   "Genome sequence of Halobacterium species NRC-1.";
RL   Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).
DR   EMBL; AE004978; AAG18714.1; -.
KW   Complete proteome.
SQ   SEQUENCE   71 AA;  7826 MW;  [illegible] CRC64;

  Query Match             100.0%;  Score 6;  DB 17;  Length 71;
  Best Local Similarity   50.0%;  Pred. No. 3.7e+03;
  Matches    3;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 WXXWXF 6
          |::|:|
Db     49 WWILWGF 54


RESULT 12
Q50168
ID   Q50168      PRELIMINARY;      PRT;    74 AA.
AC   Q50168;
DT   01 NOV 1996 (TrEMBLrel. 01, Created)
DT   01 NOV 1996 (TrEMBLrel. 01, Last sequence update)
DT   01 NOV 1996 (TrEMBLrel. 01, Last annotation update)
DE   U296u.
OS   Mycobacterium leprae.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1769;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Smith D.R.;
RL   Submitted (AUG 1996) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Robison K.;
RL   Submitted (SEP 1996) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U15187; AAA63114.1; -.
SQ   SEQUENCE   74 AA;  8343 MW;  [illegible] CRC64;

  Query Match             100.0%;  Score 6;  DB 2;  Length 74;
  Best Local Similarity   50.0%;  Pred. No. 3.8e+03;
  Matches    3;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 WXXWXF 6
          |::|:|
Db     59 WVKWLF 64


RESULT 13
```

```
O68793
ID   O68793      PRELIMINARY;      PRT;    76 AA.
AC   O68793;
DT   01-AUG-1998 (TrEMBLrel. 07, Created)
DT   01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Hypothetical protein Y1091.
GN   Y1091 OR YPMT1.60C.
OS   Yersinia pestis.
OG   Plasmid pMT1 (pMT-1).
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Yersinia.
OX   NCBI_TaxID=632;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KIM5;
RX   MEDLINE=98422474; PubMed=9748454;
RA   Hu P., Elliott J., McCready P., Skowronski E., Garnes J.,
RA   Kobayashi A., Brubaker R.R., Garcia E.;
RT   "Structural organization of virulence-associated plasmids of Yersinia
RT   pestis.";
RL   J. Bacteriol. 180:5192-5202(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KIM10+;
RX   MEDLINE=99043898; PubMed=9826348;
RA   Lindler L.E., Plano G.V., Burland V., Mayhew G.F., Blattner F.R.;
RT   "Complete DNA sequence and detailed analysis of the Yersinia pestis
RT   KIM5 plasmid encoding murine toxin and capsular antigen.";
RL   Infect. Immun. 66:5731-5742(1998).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CO-92 / BIOVAR ORIENTALIS; PLASMID=PMT1 (PMT 1);
RX   MEDLINE=21470413; PubMed=11586360;
RA   Parkhill J., Wren B.W., Thomson N.R., Titball R.W., Holden M.T.G.,
RA   Prentice M.B., Sebaihia M., James K.D., Churcher C., Mungall K.L.,
RA   Baker S., Basham D., Bentley S.D., Brooks K., Cerdeno-Tarraga A.M.,
RA   Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,
RA   Feltwell T., Hamlin N., Holroyd S., Jagels K., Karlyshev A.V.,
RA   Leather S., Moule S., Oyston P.C.F., Quail M., Rutherford K.,
RA   Simmonds M., Skelton J., Stevens K., Whitehead S., Barrell B.G.;
RT   "Genome sequence of Yersinia pestis, the causative agent of plague.";
RL   Nature 413:523-527(2001).
DR   EMBL; AF053947; AAC13243.1; -.
DR   EMBL; AF074611; AAC82749.1; -.
DR   EMBL; AL117211; CAB55242.1; -.
KW   Hypothetical protein; Plasmid; Complete proteome.
SQ   SEQUENCE   76 AA;  8807 MW;  7EF44937755036BB CRC64;

  Query Match             100.0%;  Score 6;  DB 16;  Length 76;
  Best Local Similarity    50.0%;  Pred. No. 3.9e+03;
  Matches    3;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 WXXWXF 6
           |::|:|
Db      50 WRTWNF 55


RESULT 14
Q9S4W3
ID   Q9S4W3      PRELIMINARY;      PRT;    77 AA.
AC   Q9S4W3;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   Hypothetical 9.1 kDa protein.
OS   Escherichia coli.
OG   Plasmid F.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
OX   NCBI_TaxID=562;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99206678; PubMed=10366527;
RA   Manwaring N.P., Skurray R.A., Firth N.;
RT   "Nucleotide sequence of the F plasmid leading region.";
RL   Plasmid 41:219-225(1999).
DR   EMBL; AF106329; AAD47188.1; -.
KW   Hypothetical protein; Plasmid.
SQ   SEQUENCE   77 AA;  9083 MW;  36745FF0267B977A CRC64;

  Query Match             100.0%;  Score 6;  DB 2;  Length 77;
  Best Local Similarity    50.0%;  Pred. No. 4e+03;
  Matches    3;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 WXXWXF 6
           |::|:|
Db      53 WRTWNF 58


RESULT 15
Q9WTF7
ID   Q9WTF7      PRELIMINARY;      PRT;    79 AA.
AC   Q9WTF7;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   YdeA protein.
GN   YDEA.
OS   Escherichia coli.
OG   Plasmid P100.
OC   Bacteria, Proteobacteria, gamma subdivision, Enterobacteriaceae;
OC   Escherichia.
OX   NCBI_TaxID=562;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Sampei G., Mizobuchi K.;
RT   "Organization and diversification of plasmid genomes: complete
RT   nucleotide sequence of the P100 genome.";
RL   Submitted (MAY-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AP000342; BAA78827.1; -.
SQ   SEQUENCE   79 AA;  9298 MW;  71CFB485FFFC1A6C CRC64;

  Query Match             100.0%;  Score 6;  DB 2;  Length 79;
  Best Local Similarity    50.0%;  Pred. No. 4e+03;
  Matches    3;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 WXXWXF 6
           |::|:|
Db      55 WRTWNF 60


Search completed: February 20, 2003, 11:28:57
Job time : 47 secs
```

GenCore version 5.1.3

OM protein - protein search, using sw model

Run on: February 20, 2003, 11:21:07 , Search time 29.3333 Seconds
(without alignments)
27.256 Million cell updates/sec

Title: US-09-912-414-9
Perfect score: 6
Sequence: 1 WXXWXF 6

Scoring table: OLIGODX
Gapop 60.0 , Gapext 60.0

Searched: 908470 seqs, 133250620 residues

Word size: 0

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database: A_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	6	100.0	6	15	AAR57386	Peptide for treati
2	6	100.0	6	15	AAR57391	Peptide for treati
3	6	100.0	6	21	AAB01492	Peptide which bind
4	6	100.0	6	21	AAB01497	Peptide which bind
5	6	100.0	6	21	AAB01499	Peptide which bind
6	6	100.0	6	21	AAB01505	Peptide which bind
7	6	100.0	6	21	AAB01506	Peptide which bind
8	6	100.0	6	21	AAB01508	Peptide which bind
9	6	100.0	7	22	AAE12181	Polyglutamine repe
10	6	100.0	7	22	AAM45777	H11 binding site c
11	6	100.0	8	15	AAR57399	Peptide for treati
12	6	100.0	9	21	AAB01498	Peptide which bind
13	6	100.0	9	22	AAU93672	Granulocyte colony
14	6	100.0	10	15	AAR57400	Peptide for treati
15	6	100.0	11	19	AAW62149	Haemophilus influe
16	6	100.0	11	21	AAB20714	Polymeric immunogl
17	6	100.0	12	15	AAR57401	Peptide for treati
18	6	100.0	12	21	AAY88104	Oregon green 514 b
19	6	100.0	12	21	AAY88154	Oregon green 514 b
20	6	100.0	13	18	AAW38112	Dystrophin WW doma
21	6	100.0	13	23	ABP46201	Human BLyS binding
22	6	100.0	14	22	AAE07760	Human HLA-DP restr
23	6	100.0	15	20	AAY30351	Epitope derived fr
24	6	100.0	15	21	AAY65509	Oestrogen receptor
25	6	100.0	15	23	AAU77026	Human ATP-dependen
26	6	100.0	15	22	AAU86245	Oestrogen receptor
27	6	100.0	16	23	AAB19239	Streptococcus pneu
28	6	100.0	17	21	AAR09193	Hepatitis GB virus
29	6	100.0	19	22	AAB76504	Peptide which bind
30	6	100.0	20	23	AAU90618	Insulin/insulin-li
31	6	100.0	20	23	AAU90940	Insulin/insulin-li
32	6	100.0	20	23	AAU70773	Hepatitis B virus
33	6	100.0	21	22	ABG04111	Novel human diagno
34	6	100.0	21	22	ABG28697	Novel human diagno
35	6	100.0	22	22	ABB79009	Consensus signal s
36	6	100.0	22	22	AAE13402	Consensus signal p
37	6	100.0	22	22	AAM52570	Consensus signal p
38	6	100.0	22	22	AAE13132	Consensus signal p
39	6	100.0	22	22	AAE12406	Consensus signal p
40	6	100.0	22	23	AAU99514	Consensus signal s
41	6	100.0	22	23	ABG63324	Synthetic signal p
42	6	100.0	22	23	AAE20951	Consensus signal p
43	6	100.0	22	23	AAU79154	Synthetic consensu
44	6	100.0	22	23	AAU75223	Consensus signal p
45	6	100.0	22	23	AAU75412	Tumour necrosis fa

ALIGNMENTS

RESULT 1
AAR57386
ID AAR57386 standard, peptide; 6 AA.
XX
AC AAR57386;
XX
DT 21-MAR-1995 (first entry)
XX
DE Peptide for treating diseases related to anti DNA antibodies.
XX
KW Carrier; absorbing agent, treatment; anti-DNA antibody, immune complex.
XX
OS Syntheric
XX
PN JP06192290-A.
XX
PD 12-JUL-1994.
XX
PF 18-JAN-1993; 93JP-0006098.
XX
PR 30-SEP-1992; 92JP-0261821.
XX
PA (KUPS) KUPAPAY CO LTD
XX
DR WPI; 1994-260510/32.
XX
PT A peptide and an adsorbing agent prepd. by immobilising it on a
PT carrier - useful for treatment of diseases related to anti DNA
PT antibodies and immune complexes
XX
PS Disclosure; Page 11; 14pp; Japanese.
XX
CC The sequences given in AAR57386-413 are peptides which are all covered

CC by the claimed generic formula.
CC H-X-(A-B)n-Y-Z
CC A = Trp, Phe or a peptide fragment consisting of 2 residues;
CC B = Trp, Phe, Asn or Glu;
CC X and Y = a bond or Asp, Glu, Arg, Lys, His or a peptide fragment
CC consisting of 2-10 residues, provided that at least one of
CC X or Y are present;
CC Z = OH or NH2; and
CC n = 2-5.
CC These peptides may be immobilised on a carrier in the preparation of an
CC absorbing agent which may be used in the treatment of diseases realated
CC to anti-DNA antibodies and/or immune compex.
XX
SQ Sequence 6 AA;

Query Match 100.0%; Score 6; DB 15; Length 6;
Best Local Similarity 100.0%; Pred. No. 7.8e+05;
Matches 3; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 WXXWXF 6
|::|:|
Db 1 WFFWFF 6

RESULT 2
AAR57391
ID AAR57391 standard; Protein; 6 AA.
XX
AC AAR57391;
XX
DT 21-MAR-1995 (first entry)
XX
DE Peptide for treating diseases related to anti-DNA antibodies.
XX
KW Carrier, absorbing agent, treatment, anti-DNA antibody, immune compex.
XX
OS Synthetic
XX
PN JP06192291-A.
XX
PD 12-JUL-1994.
XX
PF 18-JAN-1993; 93JP-0006008.
XX
PR 30-SEP-1992; 92JP-0261811.
XX
PA (KURS) KURARAY CO LTD
XX
DR WPI; 1994-261813/32.
XX
PT A peptide and an absorbing agent prepd. by immobilising it on a
PT carrier - useful for treatment of diseases related to anti-DNA
PT antibodies and immune complexes
XX
PS Disclosure; Page 11; 14pp; Japanese
XX
CC The sequences given in AAR57396-413 are peptides which are all covered
CC by the claimed generic formula:
CC H-X-(A-B)n-Y-Z
CC A = Trp, Phe or a peptide fragment consisting of 2 residues;
CC B = Trp, Phe, Asn or Glu;
CC X and Y = a bond or Asp, Glu, Arg, Lys, His or a peptide fragment
CC consisting of 2-10 residues, provided that at least one of
CC X or Y are present;
CC Z = OH or NH2; and
CC n = 2-5.
CC These peptides may be immobilised on a carrier in the preparation of an
CC absorbing agent which may be used in the treatment of diseases realated
CC to anti DNA antibodies and/or immune compex.
XX
SQ Sequence 6 AA;

Query Match 100.0%; Score 6; DB 15; Length 6;
Best Local Similarity 100.0%; Pred. No. 7.8e+05;
Matches 3; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 WXXWXF 6
|::|:|
Db 1 WFFWFF 6

RESULT 3
AAB01492
ID AAB01492 standard; peptide; 6 AA.
XX
AC AAB01492;
XX
DT 08-NOV-2000 (first entry)
XX
DE Peptide which binds to transcription factor E2F 1 DNA binding domain.
XX
KW DNA binding, transcription factor; E2F; E2F 1; cell cycle; DP-1;
KW activation; transcription; apoptosis; proliferative disorder;
KW psoriasis; restenosis.
XX
OS Synthetic.
XX
PN WO200044771-A1.
XX
PD 03-AUG-2000.
XX
PF 26-JAN-2000; 2000WO-GB00227.
XX
PR 26-JAN-1999; 99GB-0001710
XX
PA (PROL) PROLIFIX LTD.
XX
PI Mueller F, Montermann RB, Montigiani S;
XX
DR WPI; 2000-532906/49.
XX
PT Peptides binding to the DNA binding domain of transcription factor E2F
PT and inhibiting cell cycle progression, useful for the treatment of
PT cancer
XX
PS Claim 6, Page 2, 42pp; English.
XX
CC Peptides which bind to the DNA binding domain of transcription
CC factor E2F and inhibit cell cycle progression may be useful as
CC research agents to investigate the interaction between E2F and DP-1,
CC in the activation of transcription by E2F 1/DP 1 heterodimers. They
CC may also be used for inducing apoptosis and/or cell cycle arrest in
CC a cell, particularly for treatment of cancer or other proliferative
CC disorders such as psoriasis and restenosis.
XX
SQ Sequence 6 AA;

Query Match 100.0%; Score 6; DB 21; Length 6;
Best Local Similarity 50.0%; Pred. No. 7.8e+05;
Matches 3; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 WXXWXF 6
|::|:|
Db 1 WVRWHF 6

RESULT 4
AAB01497
ID AAB01497 standard; peptide; 6 AA.
XX
AC AAB01497;
XX
DT 08-NOV-2000 (first entry)
XX
DE Peptide which binds to transcription factor E2F 1 DNA binding domain.
XX

```
KW   DNA binding, transcription factor; E2F, E2F-1; cell cycle, DP 1;
KW   activation; transcription, apoptosis, proliferative disorder,
KW   psoriasis; restenosis.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 2
FT                   /note= "Any amino acid"
FT   Misc-difference 3
FT                   /note= "Any amino acid"
FT   Misc-difference 5
FT                   /note= "Any amino acid"
XX
PN   WO200044771-A1.
XX
PD   03-AUG-2000.
XX
PF   26-JAN-2000; 2000WO-GB00227.
XX
PR   26-JAN-1999;   99GB-0001710.
XX
PA   (PROL-) PROLIFIX LTD.
XX
PI   Mueller R,  Kontermann RE,  Montigiani S;
XX
DR   WPI; 2000-532806/48.
XX
PT   Peptides binding to the DNA binding domain of transcription factor E2F
PT   and inhibiting cell cycle progression, useful for the treatment of
PT   cancer
XX
PS   Claim 4; Page 9; 42pp; English.
XX
CC   Peptides which bind to the DNA binding domain of transcription
CC   factor E2F and inhibit cell cycle progression may be useful as
CC   research agents to investigate the interaction between E2F and DP 1,
CC   or the activation of transcription by E2F-1/DP-1 heterodimers. They
CC   may also be used for inducing apoptosis and/or cell cycle arrest in
CC   a cell, particularly for treatment of cancer or other proliferative
CC   disorders such as psoriasis and restenosis.
XX
SQ   Sequence   6 AA;

  Query Match           100.0%;  Score 6;  DB 21;  Length 6;
  Best Local Similarity  100.0%;  Pred. No. 7.8e-05;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 WXXWXF 6
           ||||||
Db       1 WXXWXF 6


RESULT 5
AAB01499
ID   AAB01499 standard; peptide; 6 AA.
XX
AC   AAB01499;
XX
DT   08-NOV-2000  (first entry)
XX
DE   Peptide which binds to transcription factor E2F-1 DNA binding domain.
XX
KW   DNA binding; transcription factor; E2F; E2F-1; cell cycle; DP-1;
KW   activation; transcription; apoptosis; proliferative disorder;
KW   psoriasis; restenosis.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 2
FT                   /note= "Any amino acid"
FT   Misc-difference 3
FT                   /note= "Any amino acid"
XX
PN   WO200044771-A1.
XX
PD   03-AUG-2000.
XX
PF   26-JAN-2000; 2000WO-GB00227.
XX
PR   26-JAN-1999;   99GB-0001710.
XX
PA   (PROL-) PROLIFIX LTD.
XX
PI   Mueller R,  Kontermann RE,  Montigiani S;
XX
DR   WPI; 2000-532806/48.
XX
PT   Peptides binding to the DNA binding domain of transcription factor E2F
PT   and inhibiting cell cycle progression, useful for the treatment of
PT   cancer
XX
PS   Claim 4; Page 9; 42pp; English.
XX
CC   Peptides which bind to the DNA binding domain of transcription
CC   factor E2F and inhibit cell cycle progression may be useful as
CC   research agents to investigate the interaction between E2F and DP-1,
CC   or the activation of transcription by E2F-1/DP-1 heterodimers. They
CC   may also be used for inducing apoptosis and/or cell cycle arrest in
CC   a cell, particularly for treatment of cancer or other proliferative
CC   disorders such as psoriasis and restenosis.
XX
SQ   Sequence   6 AA;

  Query Match           100.0%;  Score 6;  DB 21;  Length 6;
  Best Local Similarity  83.3%;  Pred. No. 7.8e-05;
  Matches    5;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 WXXWXF 6
           ||||:|
Db       1 WXXWHF 6


RESULT 6
AAB01505
ID   AAB01505 standard; peptide; 6 AA.
XX
AC   AAB01505;
XX
DT   08-NOV-2000  (first entry)
XX
DE   Peptide which binds to transcription factor E2F-1 DNA binding domain.
XX
KW   DNA binding; transcription factor; E2F; E2F-1; cell cycle; DP-1;
KW   activation; transcription; apoptosis; proliferative disorder;
KW   psoriasis; restenosis.
XX
OS   Synthetic.
XX
PN   WO200044771-A1.
XX
PD   03-AUG-2000.
XX
PF   26-JAN-2000; 2000WO-GB00227.
XX
PR   26-JAN-1999;   99GB-0001710.
XX
PA   (PROL-) PROLIFIX LTD.
XX
PI   Mueller R,  Kontermann RE,  Montigiani S;
XX
DR   WPI; 2000-532806/48.
XX
PT   Peptides binding to the DNA binding domain of transcription factor E2F
PT   and inhibiting cell cycle progression, useful for the treatment of
```

```
PT   cancer
XX
PS   Example; Page 26; 42pp; English.
XX
CC   Peptides which bind to the DNA binding domain of transcription
CC   factor E2F and inhibit cell cycle progression may be useful as
CC   research agents to investigate the interaction between E2F and DP-1,
CC   or the activation of transcription by E2F 1/DP 1 heterodimers. They
CC   may also be used for inducing apoptosis and/or cell cycle arrest in
CC   a cell, particularly for treatment of cancer or other proliferative
CC   disorders such as psoriasis and restenosis.
XX
SQ   Sequence   6 AA;

  Query Match           100.0%;  Score 6;  DB 21;  Length 6;
  Best Local Similarity   50.0%;  Pred. No. 7.8e+05;
  Matches    3;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 WXXWXF 6
          |::|:|
Db      1 WARWHF 6


RESULT 7
AAB01506
ID   AAB01506 standard; peptide; 6 AA
XX
AC   AAB01506;
XX
DT   08-NOV-2000  (first entry)
XX
DE   Peptide which binds to transcription factor E2F 1 DNA binding domain.
XX
KW   DNA binding, transcription factor; E2F; E2F-1; cell cycle; DP 1;
KW   activation; transcription; apoptosis; proliferative disorder;
KW   psoriasis, restenosis.
XX
OS   Synthetic.
XX
PN   WO200044771-A1.
XX
PD   03-AUG-2000
XX
PF   26-JAN-2000; 2000WO-GB00227.
XX
PR   26-JAN 1999;   99GB-0001710.
XX
PA   (PROL-) PROLIFIX LTD.
XX
PI   Mueller R,  Kontermann RE,  Montigiani S;
XX
DR   WPI; 2000-520806/48.
XX
PT   Peptides binding to the DNA binding domain of transcription factor E2F
PT   and inhibiting cell cycle progression, useful for the treatment of
PT   cancer
XX
PS   Example; Page 26; 42pp; English.
XX
CC   Peptides which bind to the DNA binding domain of transcription
CC   factor E2F and inhibit cell cycle progression may be useful as
CC   research agents to investigate the interaction between E2F and DP 1,
CC   or the activation of transcription by E2F 1/DP 1 heterodimers. They
CC   may also be used for inducing apoptosis and/or cell cycle arrest in
CC   a cell, particularly for treatment of cancer or other proliferative
CC   disorders such as psoriasis and restenosis.
XX
SQ   Sequence   6 AA;

  Query Match           100.0%;  Score 6;  DB 21;  Length 6;
  Best Local Similarity   50.0%;  Pred. No. 7.8e+05;
  Matches    3;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 WXXWXF 6
          |::|:|
Db      1 WVAWHF 6


RESULT 8
AAB01508
ID   AAB01508 standard; peptide; 6 AA.
XX
AC   AAB01508;
XX
DT   08-NOV-2000  (first entry)
XX
DE   Peptide which binds to transcription factor E2F-1 DNA binding domain.
XX
KW   DNA binding, transcription factor, E2F, E2F 1, cell cycle; DP-1;
KW   activation, transcription, apoptosis; proliferative disorder;
KW   psoriasis; restenosis.
XX
OS   Synthetic.
XX
PN   WO200044771-A1.
XX
PD   03-AUG-2000.
XX
PF   26 JAN 2000; 2000WO-GB00227.
XX
PR   26 JAN 1999;   99GB-0001710.
XX
PA   (PROL-) PROLIFIX LTD.
XX
PI   Mueller R,  Kontermann RE,  Montigiani S;
XX
DR   WPI; 2000-520806/48.
XX
PT   Peptides binding to the DNA binding domain of transcription factor E2F
PT   and inhibiting cell cycle progression, useful for the treatment of
PT   cancer
XX
PS   Example; Page 26; 42pp; English.
XX
CC   Peptides which bind to the DNA binding domain of transcription
CC   factor E2F and inhibit cell cycle progression may be useful as
CC   research agents to investigate the interaction between E2F and DP-1,
CC   or the activation of transcription by E2F 1/DP 1 heterodimers. They
CC   may also be used for inducing apoptosis and/or cell cycle arrest in
CC   a cell, particularly for treatment of cancer or other proliferative
CC   disorders such as psoriasis and restenosis.
XX
SQ   Sequence   6 AA;

  Query Match           100.0%;  Score 6;  DB 21;  Length 6;
  Best Local Similarity   50.0%;  Pred. No. 7.8e+05;
  Matches    3;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 WXXWXF 6
          |::|:|
Db      1 WVRWAF 6


RESULT 9
AAE12181
ID   AAE12181 standard; peptide; 7 AA.
XX
AC   AAE12181;
XX
DT   18-DEC-2001  (first entry)
XX
DE   Polyglutamine repeat domain binding peptide, QPB1-M7FD.
XX
KW   Polyglutamine repeat domain, huntingtin, atrophin, ataxin;
KW   [illegible] neurodegenerative disease, antiparkinsonian;
KW   Huntington's disease, dentatorubral pallidoluysian atrophy; DRPLA;
```

KW spinobulbar muscular atrophy; spinocerebellar ataxia; nootropic
XX
OS Unidentified.
XX
PN WO200168678-A2.
XX
PD 20-SEP-2001.
XX
PF 14-MAR-2001; 2001WO-US08222.
XX
PR 16-MAR-2000; 2000US-189781P.
XX
PA (UYDU) UNIV DUKE.
XX
PI Strittmatter WJ, Burke JR, Nagai Y;
XX
DR WPI; 2001-616286/71.
XX
PT New polypeptides which selectively bind to expanded polyglutamine
PT repeat domains, useful for treating neurodegenerative diseases, e.g.
PT Huntington's disease -
XX
PS Example 13; Page 14; 59pp; English.
XX
CC The present sequence is a polyglutamine repeat domain binding peptide.
CC The peptide is useful for treating a cell that contains and expresses a
CC protein having an expanded polyglutamine region (e.g. huntingtin;
CC atrophin 1, ataxin 1, 2, 6 or 7; or androgen receptor protein), or
CC treating a neurodegenerative disease characterised by the presence of
CC expanded polyglutamine repeats, e.g. Huntington's disease, dentatorubral
CC pallidoluysian atrophy (DRPLA), spinobulbar muscular atrophy, and
CC spinocerebellar ataxia types 1, 2, 3, 6 and 7
XX
SQ Sequence 7 AA;

Query Match 100.0%; Score 6; DB 22; Length 7;
Best Local Similarity 50.0%; Pred. No. 7.8e+05;
Matches 3; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 WXXWXF 6
|::|:|
Db 1 WKWWPF 6

RESULT 10
AAM45777
ID AAM45777 standard; Peptide; 7 AA.
XX
AC AAM45777;
XX
DT 25-OCT-2001 (first entry)
XX
DE H11 binding site consensus conforming peptide (CCP) #2048.
XX
KW Antigen-binding, tumour, diagnosis, stress protein-peptide complex, SPPC,
KW immunogenically cross-reactive; cancer; immunogenic cancer cell;
KW cytostatic; vaccine; tumour specific immunogenic response inducer,
KW astrocytoma, fibrosarcoma, myxosarcoma, liposarcoma, oligodendroglioma,
KW ependymoma, medulloblastoma, primitive neural ectodermal tumour.
XX
OS Homo sapiens
OS Synthetic.
XX
PN CA2290722-A1.
XX
PD 08-JUN-2001.
XX
PF 08-DEC-1999; 99CA-2290722.
XX
PR 08-DEC-1999; 99CA-2290722.
XX
PA (NOVO-) NOVOPHARM BIOTECH INC.
XX
PI Kaplan HA, Maiti PK, Fast DG, Herman W, Dan MD, Lewis KE;
PI Entwistle JM, MacDonald GC;
XX
DR WPI; 2001-425937/46.
XX
PT Composition useful for treating and diagnosing cancer, comprises stress
PT protein-peptide complexes associated with tumor, and isolated
PT antigen-binding fragments of an antibody that binds specifically to the
PT complex -
XX
PS Example 4; Page 108; 154pp; English.
XX
CC The present invention describes a composition (I) comprising stress
CC protein-peptide complexes (SPPC) associated with tumours that is
CC specifically immunogenically cross-reactive with cell surface associated
CC SPPCs specific to target cancer (TC). Also described is an isolated
CC antigen-binding fragment of an antibody that binds specifically to SPPCs
CC or a population of different SPPCs consisting of immunogenic cancer cell
CC surface-associated SPPC of TC. (I) has cytostatic activity and can be
CC used in vaccine production and as a tumour-specific immunogenic response
CC inducer. (I) is useful for treating 71 types of cancers or tumours in a
CC subject, such as astrocytoma, fibrosarcoma, myxosarcoma, liposarcoma,
CC oligodendroglioma, ependymoma, medulloblastoma, and primitive neural
CC ectodermal tumour (PNET). (I) is useful as cancer immunogen including
CC vaccines. (I) is useful for diagnostic and palliative use, for detecting
CC or imaging cancer cells, and to monitor the course of amelioration of
CC malignancy in an individual. AAM47107 to AAM47109 represent peptides
CC which are used in the exemplification of the present invention.
XX
SQ Sequence 7 AA;

Query Match 100.0%; Score 6; DB 22; Length 7;
Best Local Similarity 50.0%; Pred. No. 7.8e+05;
Matches 3; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 WXXWXF 6
|::|:|
Db 1 WRRWNF 6

RESULT 11
AAR57399
ID AAR57399 standard; Protein; 8 AA.
XX
AC AAR57399;
XX
DT 21-MAR-1995 (first entry)
XX
DE Peptide for treating diseases related to anti-DNA antibodies.
XX
KW Carrier, absorbing agent, treatment, anti DNA antibody, immune complex.
XX
OS Synthetic.
XX
PN JP06122696-A.
XX
PD 12-JUL-1994.
XX
PF 18-JAN-1993; 93JP-0006098.
XX
PR 30-SEP-1992; 92JP-0261821.
XX
PA (KURS) KURARAY CO LTD.
XX
DR WPI; 1994-260510/32.
XX
PT A peptide and an adsorbing agent prepd. by immobilising it on a
PT carrier useful for treatment of diseases related to anti DNA
PT antibodies and immune complexes
XX
PS Disclosure; Page 12; 14pp; Japanese.
XX
CC The sequences given in AAR57386-413 are peptides which are all covered

```
CC   by the claimed general formula
CC                        H-X-(A-B)n-Y-Z
CC   A = Trp, Phe or a peptide fragment consisting of 2 residues,
CC   B = Trp, Phe, Asn or Glu,
CC   X and Y = a bond or Asp, Glu, Arg, Lys, His or a peptide fragment
CC             consisting of 2-10 residues, provided that at least one of
CC             X or Y are present;
CC   Z = OH or NH2, and
CC   n = 2-5.
CC   These peptides may be formulated in a carrier in the preparation of an
CC   absorbing agent which may be used in the treatment of diseases realated
CC   to anti DNA antibodies and a immune complex
XX
SQ   Sequence   8 AA;

  Query Match             100.0%;  Score 6;  DB 15;  Length 8;
  Best Local Similarity   50.0%;  Pred. No. 7.8e+05;
  Matches    3;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 WXXWXF 6
           |:::|:|
Db       1 WWFWWF 6


RESULT 10
AAB01498
ID   AAB01498 standard; peptide; 9 AA
XX
AC   AAB01498;
XX
DT   08-NOV 2000  (first entry)
XX
DE   Peptide which binds to transcription factor E2F 1 DNA binding domain.
XX
KW   DNA binding, transcription factor; E2F; E2F 1; cell cycle; DP 1;
KW   activation, transcription, apoptosis, proliferative disorder,
KW   psoriasis; restenosis.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 2
FT                   /note= "Any amino acid"
FT   Misc-difference 3
FT                   /note= "Any amino acid"
FT   Misc-difference 5
FT                   /note= "Any amino acid"
FT   Misc-difference 7
FT                   /note= "Any amino acid"
FT   Misc-difference 8
FT                   /note= "Any amino acid"
XX
PN   WO200014771 A1.
XX
PD   03-AUG-2000
XX
PF   26-JAN-2000; 2000WO-GB00227.
XX
PR   26-JAN-1999;  1999GB-000001618
XX
PA   (PROL-) PROLIFIX LTD.
XX
PI   Mueller R, Knuttermann RE, Montigiani S;
XX
DR   WPI; 2000-532806/48.
XX
PT   Peptides binding to the DNA binding domain of transcription factor E2F
PT   and inhibiting cell cycle progression, useful for the treatment of
PT   cancer
XX
PS   Claim 4; Page 9; 42pp; English.
XX
CC   Peptides which bind to the DNA binding domain of transcription
CC   factor E2F and inhibit cell cycle progression may be useful as
CC   research agents to investigate the interaction between E2F and DP-1,
CC   or the activation of transcription by E2F 1/DP-1 heterodimers. They
CC   may also be used for inducing apoptosis and/or cell cycle arrest in
CC   a cell, particularly for treatment of cancer or other proliferative
CC   disorders such as psoriasis and restenosis.
XX
SQ   Sequence   9 AA;

  Query Match             100.0%;  Score 6;  DB 21;  Length 9;
  Best Local Similarity   100.0%;  Pred. No. 7.8e+05;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 WXXWXF 6
           ||||||
Db       1 WXXWXF 6


RESULT 13
AAU93672
ID   AAU93672 standard; Peptide; 9 AA.
XX
AC   AAU93672;
XX
DT   02 JUL 2002  (first entry)
XX
DE   Granulocyte colony stimulating factor receptor binding peptide #478.
XX
KW   G CSFR, granulocyte colony stimulating factor receptor, cytokine;
KW   haematopoietic growth factor; neutrophil proliferation; AIDS;
KW   neutrophil differentiation, acquired immunodeficiency syndrome;
KW   chemotherapy-induced neutropaenia; community acquired pneumonia;
KW   depressed neutrophil count, immunostimulant.
XX
OS   Synthetic.
XX
PN   WO200207676 A2.
XX
PD   31-JAN-2002.
XX
PF   20 JUL 2001; 2001WO-US23046.
XX
PR   20 JUL-2000; 2000US-0620091.
XX
PA   (GLAX ) GLAXO GROUP LTD.
XX
PI   Cwirla SE, Balu P, Duffin DJ, Piplani S, McBowen-Merrill B;
PI   Schatz PJ;
XX
DR   WPI; 2002-329382/36.
XX
PT   Novel compounds, useful for treating depressed neutrophil count,
PT   comprise peptide chains of approximately 6 to 40 amino acids in length
PT   that bind to granulocyte colony stimulating factor receptor
XX
PS   Claim 41; Page 67; 90pp; English.
XX
CC   The invention relates to compounds comprising a peptide chain
CC   approximately 6 to 40 amino acids in length that binds to granulocyte-
CC   colony stimulating factor receptor (G-CSFR). The compounds contain
CC   specific sequences of the generic peptides appearing as AAU79400-AAU79409
CC   and the generic sequences XV_1XV_2XV_3XV_4XV_5XV_6XV_7XV_8 (where
CC   XV_1 = E, C, Q, V or Y; XV_2 = P, A, I, M, S, W or Q; XV_3 = K, P or T;
CC   XV_4 = L, A or Y, XV_5 = R, A, M, H, E, V, L, G, D, Q or S; XV_6 = E or
CC   Y, XV_7 = A or G, and XV_8 = P, H, G or D) and XVI_1XVI_2XVI_3XVI_4XVI_5
CC   EXVI_6XVI_7XVI_8XVI_9 (where XVI_1 = A, E or G, XVI_2 = E, H or D;
CC   XVI_3 = R or G, XVI_4 = R, Y, M, N, G, P, D, I, S or E, XVI_5 = A, S or
CC   P; XVI_6 = E, D, T, Q, H or A, XVI_7 = E, W, H, L, G, A or Q, XVI_8 = P
CC   or E, and XVI_9 = W, G or R). The compounds are used for treating
CC   conditions associated with depressed neutrophil count e.g. chemotherapy-
CC   induced neutropaenia, AIDS-induced neutropaenia or community-acquired
CC   pneumonia-induced pneumonia. The compounds are useful as in vitro as
CC   tools for understanding the biological role of granulocyte-colony
```

```
CC   stimulating factor (G-CSF a haematopoietic growth factor and
CC   cytokine that stimulates neutrophil proliferation and differentiation),
CC   including evaluation of many factors thought to influence, and be
CC   influenced by, production of white blood cells, in the development of
CC   compounds that bind to G-CSFR, as reagents for detecting G-CSF receptor
CC   or related receptor on living cells, fixed cells, in biological fluid, in
CC   tissue homogenates or in purified natural biological materials, in situ
CC   staining, fluorescence-activated cell sorting (FACS), Western blotting or
CC   enzyme-linked immunoadsorptive assay (ELISA), in receptor purification or
CC   in purifying cells expressing G-CSFR on the cell surface (or inside
CC   permeabilised cells) as a commercial research reagent for various medical
CC   and diagnostic uses or to treat a disease that would benefit from the
CC   ability to of a compound to mimic the effects of G-CSF in vivo
CC   The compounds bind specifically to G-CSFR and allow for studies of
CC   biological activities mediated by the receptor and for the treatment of
CC   diseases, disorders and conditions that would benefit from activating or
CC   inactivating G-CSFR. The present sequence is a G-CSFR binding peptide of
CC   the invention.
XX
SQ   Sequence   9 AA;

  Query Match             100.0%;  Score 6;  DB 23;  Length 9;
  Best Local Similarity   50.0%;  Pred. No. 7.9e+05;
  Matches    3;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 WXXWXF 6
          |::|:|
Db      4 WWGWNF 9


RESULT 14
AAR57400
ID   AAR57400 standard; Protein; 10 AA.
XX
AC   AAR57400;
XX
DT   21-MAR-1995  (first entry)
XX
DE   Peptide for treating diseases related to anti-DNA antibodies
XX
KW   Carrier; absorbing agent; treatment; anti-DNA antibody; immune complex
XX
OS   Synthetic
XX
PN   JP06192290-A.
XX
PD   12-JUL-1994.
XX
PF   18-JAN-1993;   93JP-0006098.
XX
PR   30-SEP-1992;   92JP-0261921.
XX
PA   (KUPS ) KUPARAY CO LTD.
XX
DR   WPI; 1994-260510/32.
XX
PT   A peptide and an absorbing agent prepd. by immobilising it on a
PT   carrier - useful for treatment of diseases related to anti-DNA
PT   antibodies and immune complexes
XX
PS   Disclosure; Page 12; 14pp; Japanese.
XX
CC   The sequences given in AAR57386-413 are peptides which are all covered
CC   by the claimed generic formula:
CC                          H-X-(A-B)n-Y-Z
CC   A = Trp, Phe or a peptide fragment consisting of 2 residues;
CC   B = Trp, Phe, Asn or Glu;
CC   X and Y = a bond or Asp, Glu, Arg, Lys, His or a peptide fragment
CC             consisting of 2-10 residues, privided that at least one of
CC             X or Y are present;
CC   Z = OH or NH2; and
CC   n = 2-5.
CC   These peptides may be immobilised on a carrier in the preparation of an
CC   absorbing agent which may be used in the treatment of diseases realated
CC   to anti-DNA antibodies and/or immune compex.
XX
SQ   Sequence   10 AA;

  Query Match             100.0%;  Score 6;  DB 15;  Length 10;
  Best Local Similarity   50.0%;  Pred. No. 9.5e+02;
  Matches    3;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 WXXWXF 6
          |..|.|
Db      3 WWFWWF 8


RESULT 15
AAW62149
ID   AAW62149 standard; peptide; 11 AA.
XX
AC   AAW62149;
XX
DT   16-SEP-1998  (first entry)
XX
DE   Haemophilus influenzae tyrosine tRNA synthetase binding peptide 22.
XX
KW   Identification; ligand; biological activity; target-binding;
KW   drug screening; library; inhibitory ligand.
XX
OS   Synthetic.
OS   Haemophilus influenzae.
XX
PN   WO9819162-A1.
XX
PD   07-MAY-1998.
XX
PF   31-OCT-1997;   97WO-US19638.
XX
PR   31-OCT-1996;   96US-0740671.
XX
PA   (NOVA-) NOVALON PHARM CORP.
XX
PI   Fowlkes DM,  Fielinger JA,  Hyde-Deruyscher RP,  Kay BK;
XX
DR   WPI; 1998-272389/24.
XX
PT   Identifying ligands which mediate biological activity of a protein
PT   by identifying target-binding ligands and screening a library for
PT   ligands which inhibit target-binding ligand mediated activity
XX
PS   Example 5; Page 100; 143pp; English.
XX
CC   A method has been developed for identifying a ligand which mediates the
CC   biological activity of a target protein (T) by inhibiting the binding
CC   of (T) to a binding partner. The method comprises: (a) screening a first
CC   combinatorial library comprising first member ligands for binding to the
CC   target-binding ligands (TBLs), to identifying one or more TBLs; (b)
CC   screening a second library comprising second member ligands for the
CC   ability to inhibit the binding of one or more of the TBLs to the target
CC   protein, and so obtaining one or more inhibitory ligands; and (c)
CC   determining which of the inhibitory ligands can mediate a biological
CC   activity of the target protein. The present sequence represents a
CC   potential binding peptide for Haemophilus influenzae tyrosine tRNA
CC   synthetase from an example of the present invention. The method can be
CC   used for identifying drugs which can mediate the biological activity of
CC   a target protein. It can be used to identify the biological activity of
CC   a target protein whose biological function is not known and perhaps
CC   cannot be determined directly. The method can also be used to identify
CC   new inhibitory ligands of specific target proteins. The method provides
CC   high throughput screens which are essentially identical for similar and
CC   dissimilar targets, bypassing the need to develop distinct assays for
CC   biochemically diverse targets.
XX
SQ   Sequence   11 AA;
```

Query Match 100.0%; Score 6; DB 19; Length 11;
Best Local Similarity 50.0%; Pred. No. 9.2e+02;
Matches 3; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 WXXWXF 6
             :::::
Db       5 WPDWGF 10
```

Search completed: February 20, 2003, 11:26:32
Job time : 30.3333 secs

GenCore version 5.1.3

OM protein - protein search, using sw model

Run on: February 20, 2003, 10:46:10 ; Search time 12 Seconds
(without alignments)
48.067 Million cell updates/sec

Title: US-09-912-414-11
Perfect score: 38
Sequence: 1 WXXWHF 6

Scoring table: BLOSUM62DX
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 90134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_73:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	38	100.0	230	2	D70847	hypothetical prote
2	38	100.0	241	2	A71020	hypothetical prote
3	38	100.0	260	2	P71353	probable 2,3-dihi
4	38	100.0	293	2	A83299	hypothetical prote
5	38	100.0	358	1	S43876	site-specific DNA
6	38	100.0	358	2	A87296	modification methy
7	38	100.0	386	2	A97456	adenine-specific m
8	38	100.0	386	2	AD2674	adenine DNA methyl
9	38	100.0	394	2	D90105	putative SAR DNA-b
10	38	100.0	403	2	AF3432	site-specific DNA-
11	38	100.0	423	2	AH1407	PTS system galacti
12	38	100.0	423	2	AH1783	PTS system galacti
13	38	100.0	424	2	JC5891	omega-6 desaturase
14	38	100.0	424	2	T07742	omega-6 desaturase
15	38	100.0	439	2	T01807	hypothetical prote
16	38	100.0	443	2	T08136	probable omega-6 d
17	38	100.0	447	2	T17299	hypothetical prote
18	38	100.0	448	2	D85362	hypothetical prote
19	38	100.0	452	2	H91172	probable phosphotr
20	38	100.0	462	2	AH0964	probable PTS syste
21	38	100.0	462	2	H86018	hypothetical prote
22	38	100.0	472	2	AG0417	hypothetical prote
23	38	100.0	475	2	B84171	hypothetical prote
24	38	100.0	487	2	T32941	hypothetical prote
25	38	100.0	488	2	B87540	hypothetical prote
26	38	100.0	492	2	AH2079	alpha-amylase [imp
27	38	100.0	497	2	T40586	nucleolar protein
28	38	100.0	504	2	S48550	hypothetical prote
29	38	100.0	508	2	G00372	4-hydroxyphenylace
30	38	100.0	508	2	T50180	nucleolar protein
31	38	100.0	511	2	S58322	nucleolar protein
32	38	100.0	522	2	S96602	nucleolar protein
33	38	100.0	550	2	T06379	SAR DNA-binding pr
34	38	100.0	560	2	T06377	SAR DNA-binding pr
35	38	100.0	640	2	T28631	Y4cD protein - Rhi
36	38	100.0	718	2	T29448	hypothetical prote
37	38	100.0	729	2	AG1078	hypothetical prote
38	35	92.1	53	2	F89871	hypothetical prote
39	35	92.1	81	2	S41590	lysozyme (EC 3.2.1
40	35	92.1	139	2	S41579	lysozyme (EC 3.2.1
41	35	92.1	140	2	S39014	lysozyme (EC 3.2.1
42	35	92.1	140	2	S41573	lysozyme (EC 3.2.1
43	35	92.1	140	2	S41574	lysozyme (EC 3.2.1
44	35	92.1	140	2	S41577	lysozyme (EC 3.2.1
45	35	92.1	244	2	T01701	hypothetical prote

ALIGNMENTS

RESULT 1
D70847
hypothetical protein Rv0059 - Mycobacterium tuberculosis (strain H37Rv)
C;Species: Mycobacterium tuberculosis
C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 28-Jun-2002
C;Accession: D70847
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S
; Connor, F.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A;Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A;Reference number: A70500; MUID:98295987; PMID:9634230
A;Accession: D70847
A;Status: preliminary; nucleic acid sequence not shown, translation not shown
A;Molecule type: DNA
A;Residues: 1-230 <COL>
A;Cross-references: GB:AL021428; GB:AL123456; NID:g3261714; PIDN:CAA15240.1; PID:g3808716
A;Experimental source: strain H37Rv
C;Genetics:
A;Gene: Rv0059
C;Superfamily: Mycobacterium tuberculosis hypothetical protein Rv0059

Query Match 100.0%; Score 38; DB 2; Length 230;
Best Local Similarity 66.7%; Pred. No. 1.1e+02;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 WXXWHF 6
           |::|||
Db      26 WIVWHF 31
```

RESULT 2
A71020
hypothetical protein PH1454 - Pyrococcus horikoshii
C;Species: Pyrococcus horikoshii
C;Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 20-Jul-2000
C;Accession: A71020
R;Kawarabayasi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; Sekine
M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi,
DNA Res. 5, 55-76, 1998
A;Title: Complete sequence and gene organization of the genome of a hyper-thermophilic ar
A;Reference number: A71000; MUID:98344137; PMID:9679194
A;Accession: A71020
A;Status: preliminary; nucleic acid sequence not shown, translation not shown
A;Molecule type: DNA
A;Residues: 1-241 <KAW>
A;Cross-references: GB:AP000006; NID:g3336133; PIDN:BAA29561.1; PID:g3257878
A;Experimental source: strain OT3
A;Note: this accession replaces an interim accession for a sequence replaced by GenBank
C;Genetics:
A;Gene: PH1454

```
  Query Match             [illegible]%,  Score 4,  DB 2,  Length 241,
  Best Local Similarity   66.7%,  Pred. No. 1.1e+03,
  Matches    4;  Conservative    2,  Mismatches    0,  Indels    0,  Gaps    0,

Qy      1 WXXWHF 6
          :::||
Db    193 WQTWHF 198


RESULT 3
B71353
probable D,D-carboxypeptidase - syphilis spirochete
C;Species: Treponema pallidum subsp. pallidum (syphilis spirochete)
C;Date: 24-Jul-1999 #sequence_revision 24-Jul-1999 #text_change 05-Nov-1999
C;Accession: B71353
R;Fraser, C.M.; Norris, S.J.; Weinstock, G.M.; White, O.; Sutton, G.G.; Dodson, R.; Gwi
rson, J.; Fleischmann, R.; Richardson, D.; Shvartsbeyn, A.; Kerlavage, A.R.; Nguyen, D.
they, L.; Weidman, J.; Smith, H.O.; Venter, J.C.
Science 281, 375-388, 1998
A;Title: Complete genome sequence of Treponema pallidum, the syphilis spirochete.
A;Reference number: A71250; MUID:98332770; PMID:9665876
A;Accession: B71353
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-260 <GOD>
A;Cross-references: GB:AE000993; GB:AE000993; NID.g3322478; PID:AAC65108.1; PID.g3322483
A;Experimental source: strain Nichols
C;Genetics:
A;Gene: TP0221

  Query Match             [illegible]%,  Score 4,  DB 2,  Length [illegible],
  Best Local Similarity   66.7%,  Pred. No. 1.2e+03,
  Matches    4;  Conservative    2,  Mismatches    0,  Indels    0,  Gaps    0,

Qy      1 WXXWHF 6
          :::|
Db    220 WEPWHF 225


RESULT 4
A83299
hypothetical protein PA2778 [imported] - Pseudomonas aeruginosa (strain PAO1)
C;Species: Pseudomonas aeruginosa
C;Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C;Accession: A83299
R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrener, P.; Hickey, M.J.; Br
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A;Title: Complete genome sequence of Pseudomonas aeruginosa PAO1, an opportunistic patho
A;Reference number: A82950; MUID:20437337; PMID:10984043
A;Accession: A83299
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-293 <STO>
A;Cross-references: GB:AE004488; GB:AE004488; NID.g9949660; PID.AAG05186.1; GSPDB:GN00
A;Experimental source: strain PAO1
C;Genetics:
A;Gene: PA2778

  Query Match             100.0%,  Score 39,  DB 2,  Length 293,
  Best Local Similarity   66.7%,  Pred. No. 1.2e+03,
  Matches    4;  Conservative    2,  Mismatches    0,  Indels    0,  Gaps    0,

Qy      1 WXXWHF 6
          |::|
Db    101 WPRWHF 106


RESULT 5
S43876
site-specific DNA methyltransferase (EC 2.1.1.-) - Caulobacter crescentus
C;Species: Caulobacter crescentus
C;Date: 17-Sep-1999 #sequence_revision 17-Sep-1999 #text_change 10-Sep-1999
C;Accession: S43876
R;Zweiger, G.; Marczynski, G.; Shapiro, L.
J. Mol. Biol. 235, 472-485, 1994
A;Title: A Caulobacter DNA methyltransferase that functions only in the predivisional cel
A;Reference number: S43876; MUID:94116303; PMID:8289276
A;Accession: S43876
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-358 <ZWE>
A;Cross-references: EMBL:U01032; NID.g393011; PIDN:AAA18913.1; PID:g393012
C;Superfamily: site specific DNA methyltransferase (adenine specific); HpaI
C;Keywords: methyltransferase; restriction modification system

  Query Match             100.0%,  Score 38,  DB 1,  Length 358;
  Best Local Similarity   66.7%,  Pred. No. 1.6e+02,
  Matches    4;  Conservative    2,  Mismatches    0,  Indels    0,  Gaps    0;

Qy      1 WXXWHF 6
          |::||
Db    332 WTYWHF 337


RESULT 6
A87296
modification methylase CcrMI (imported) - Caulobacter crescentus
C;Species: Caulobacter crescentus
C;Date: 30-Apr-2001 #sequence_revision 30-Apr-2001 #text_change 17-May-2002
C;Accession: A87296
R;Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.F
 A.; Daub, M.T.; Getsoy, R.T.; Ludson, R.D.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolo
n, J.; Brinkaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.
ings Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A;Title: Complete Genome Sequence of Caulobacter crescentus.
A;Reference number: A87240; MUID:21170608; PMID:11259647
A;Accession: A87296
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-358 <STO>
A;Cross-references: GB:AE006469; NID.g13421035; PIDN:AAK22365.1; GSPDB:GN00148
C;Genetics:
A;Gene: CC0378
C;Superfamily: site-specific DNA-methyltransferase (adenine-specific); HpaI

  Query Match             100.0%,  Score 39,  DB 2,  Length 358;
  Best Local Similarity   66.7%,  Pred. No. 1.6e+02,
  Matches    4;  Conservative    2,  Mismatches    0,  Indels    0,  Gaps    0;

Qy      1 WXXWHF 6
          |::|||
Db    332 WTYWHF 337


RESULT 7
A97456
adenine specific methyltransferase CMEI (modification methylase CMEI) (imported) - Agr
C;Species: Agrobacterium tumefaciens
C;Date: 20-Sep-2001 #sequence_revision 20-Sep-2001 #text_change 17-May-2002
C;Accession: A97456
R;Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman,
 A.; Liu, F.; Wallum, S.; Allikpan, M.; Slaight, S.; Smith, C.; Lappas, C.; Markelz, R.;
Science 294, 2323-2328, 2001
A;Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tume
A;Reference number: A97359; PMID:11743194
A;Accession: A97456
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-386 <KUR>
A;Cross-references: GB:AE007869; PIDN:AAK86602.1; PID:g15156772; GSPDB:GN00163
C;Genetics:
A;Gene: AGR_C_1453
A;Map position: circular chromosome
```

C;Superfamily: site specific DNA-methyltransferase (adenine-specific) HpaI

Query Match 100.0%; Score 38; DB 2; Length 386;
Best Local Similarity 66.7%; Pred. No. 1.7e+02;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 WXXWHF 6
|::|||
Db 357 WTFWHF 362

RESULT 8
AD2674
adenine DNA methyltransferase [imported] - Agrobacterium tumefaciens (strain C58, Dupont
C;Species: Agrobacterium tumefaciens
C;Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 17-May-2002
C;Accession: AD2674
R;Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Wo
erage, C.; Gillet, W.; Grant, C.; Guenthner, D.; Koryavin, T.; Levy, R.; Li, M.; Mackell
; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A;Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
ster, E.W.
A;Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58
A;Reference number: AB2577; PMID:11743193
A;Accession: AD2674
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-386 <KUR>
A;Cross-references: GB:AE008688; PIDN:AAL41810.1; PID:g17739166; GSPDB:GN00186
A;Experimental source: strain C58 (Dupont)
C;Genetics:
A;Gene: Atu0794
A;Map position: circular chromosome
C;Superfamily: site specific DNA methyltransferase (adenine-specific) HpaI

Query Match 100.0%; Score 38; DB 2; Length 386;
Best Local Similarity 66.7%; Pred. No. 1.7e+02;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 WXXWHF 6
|::|||
Db 357 WTFWHF 362

RESULT 9
D90105
putative SAP DNA binding protein 1 [imported] - Guillardia theta nucleomorph
C;Species: nucleomorph Guillardia theta
A;Note: a nucleomorph is the vestigial nucleus of a eukaryotic endosymbiont
C;Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 24-May-2001
C;Accession: D90105
R;Douglas, S.; Zauner, S.; Fraunholz, M.; Beaton, M.; Penny, S.; Deng, L.T.; Wu, X.; Rei
Nature 410, 1091-1096, 2001
A;Title: The highly reduced genome of an enslaved algal nucleus.
A;Reference number: A90082; MUID:11323671; PMID:11323671
A;Accession: D90105
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-394 <DOU>
A;Cross-references: GB:AJ010592; NID:g3769777; PIDN:CAB39493.1; GSPDB:GN00151
C;Genetics:
A;Map position: 2
A;Genome: nucleomorph
C;Keywords: nucleomorph

Query Match 100.0%; Score 38; DB 2; Length 394;
Best Local Similarity 66.7%; Pred. No. 1.7e+02;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 WXXWHF 6
|::|||
Db 169 WYSWHF 174

RESULT 10
AF3432
site-specific DNA-methyltransferase (adenine-specific) (EC 2.1.1.72) [imported] - Brucell
C;Species: Brucella melitensis
C;Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 17-May-2002
C;Accession: AF3432
R;DelVecchio, V.G.; Kapatral, V.; Redkar, R.J.; Patra, G.; Mujer, C.; Los, T.; Ivanova, N
.; Mazur, M.; Goltsman, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Letesso
Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002
A;Title: The genome sequence of the facultative intracellular pathogen Brucella melitensi
A;Reference number: AB3252; PMID:11756688
A;Accession: AF3432
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-403 <KUR>
A;Cross-references: GB:AE009491; PIDN:AAL51641.1; PID:g17984441; GSPDB:GN00190
A;Experimental source: strain 16M
C;Genetics:
A;Gene: BMEI1444
A;Map position: I
C;Superfamily: site specific DNA methyltransferase (adenine specific) HpaI
C;Keywords: methyltransferase; S-adenosylmethionine

Query Match 100.0%; Score 38; DB 2; Length 403;
Best Local Similarity 66.7%; Pred. No. 1.8e+02;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 WXXWHF 6
|::|||
Db 373 WTFWHF 378

RESULT 11
AH1407
PTS system galactitol-specific enzyme IIC component homolog lmo2665 [imported] - Listeria
C;Species: Listeria monocytogenes
C;Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 14-Dec-2001
C;Accession: AH1407
R;Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloecker,
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.;
D.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A;Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Mat
ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez Boland, J.A.; Voss, H.; Wehland,
A;Title: Comparative genomics of Listeria species.
A;Reference number: AB1077; MUID:21637239; PMID:11673669
A;Accession: AH1407
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-423 <GLA>
A;Cross-references: GB:NC_003210; PIDN:AAL98458.1; PID:g16411150; GSPDB:GN00127
A;Experimental source: strain EGD-e
C;Genetics:
A;Gene: lmo2665
C;Superfamily: phosphotransferase enzyme II galactitol specific

Query Match 100.0%; Score 38; DB 2; Length 423;
Best Local Similarity 66.7%; Pred. No. 1.8e+02;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 WXXWHF 6
|::|||
Db 126 WNYWHF 131

RESULT 12
AH1783
PTS system galactitol-specific enzyme IIC component homolog lin2814 [imported] - Listeria
C;Species: Listeria innocua
C;Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 14-Dec-2001
C;Accession: AH1783

R;Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Perales, P.; Bloecker
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.
D.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A,Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Paraphat, G.; Malouin, E.; Matthiesen, R.; Ma
ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez Boland, J.A.; Voss, H.; Wehland,
A;Title: Comparative genomics of Listeria species.
A;Reference number: Ali[illegible]; MUID:[illegible]; PMID:[illegible]
A;Accession: AH1783
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-423 <GLA>
A;Cross-references: GB:AL592022; PIDN:CAC98040.1; PID:g16415350; GSPDB:GN00170
A;Experimental source: strain Clip11262
C;Genetics:
A;Gene: lin2814
C;Superfamily: phosphotransferase enzyme II galactitol-specific

  Query Match 100.0%; Score 38; DB 2; Length 423;
  Best Local Similarity 66.7%; Pred. No. 1.6e+02;
  Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 WXXWHF 6
|::|||
Db 126 WNYWHF 131

RESULT 13
JC5891
omega-6 fatty acid desaturase (EC 1.14.99.-), chloroplast precursor - Chlamydomonas reinhardtii
C;Species: Chlamydomonas reinhardtii
C;Date: 18-Mar-1998 #sequence_revision 18-Mar-1998 #text_change 20-Jun 2000
C;Accession: JC5891
R;Sato, N.; Fujiwara, M.; Kawaguchi, A.; Tsuzuki, M.
J. Biochem. 122, 1224-1232, 1997
A;Title: Cloning of a gene for chloroplast omega-6 desaturase of a green alga, Chlamydom
A;Reference number: JC5891; MUID:98158814; PMID:9498869
A;Accession: JC5891
A;Status: nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 1-424 <SAT>
A;Cross-references: DDBJ:AB017041; NID:g2696716; PIDN:BAA23681.1; PID:g2696717
C;Comment: This enzyme catalyzes the desaturation of a monovalent fatty acids.
C;Genetics:
A;Gene: des6
C;Superfamily: omega-3 fatty acid desaturase
C;Keywords: chloroplast; oxidoreductase
F;1-40/Domain: transit peptide (chloroplast) #status predicted <TPS

  Query Match 100.0%; Score 38; DB 2; Length 424;
  Best Local Similarity 66.7%; Pred. No. 1.8e+02;
  Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 WXXWHF 6
|::|||
Db 234 WLVWHF 239

RESULT 14
T07742
omega-6 desaturase, chloroplast - soybean
C;Species: Glycine max (soybean)
C;Date: 14 May 1999 #sequence_revision 14 May 1999 #text_change 08 Oct 1999
C;Accession: T07742
R;Hitz, W.D.; Carlson, T.J.; Booth, J.R.; Kinney, A.J.; Stecca, K.L.; Yadav, N.S.
Plant Physiol. 105, 635-641, 1994
A;Title: Cloning of a higher-plant plastid omega-6 fatty acid desaturase cDNA and its ex
A;Reference number: Z16109; MUID:94345008; PMID:8066133
A;Accession: T07742
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-424 <HIT>
A;Cross-references: EMBL:L29215; NID:g459961; PIDN:AAA60158.1; PID:g459962
A;Experimental source: seed
C;Genetics:
A;Genome: nuclear
C;Superfamily: omega-3 fatty acid desaturase
C;Keywords: chloroplast

  Query Match 100.0%; Score 38; DB 2; Length 424;
  Best Local Similarity 66.7%; Pred. No. 1.8e+02;
  Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 WXXWHF 6
|::|||
Db 251 WLMWHF 256

RESULT 15
T01807
hypothetical protein A_TM021B04.12 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 14-Feb-1997 #sequence_revision 14-Feb-1997 #text_change 02-Jun-2000
C;Accession: T01807
R;Dante, M.; Wamsley, P.; Gibson, A.
submitted to the EMBL Data Library, June 1997
A;Description: The sequence of A. thaliana TM021B04.
A;Reference number: Z14440
A;Accession: T01807
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-439 <DAN>
A;Cross-references: EMBL:AB007[illegible]; NID:g[illegible]; PIDN:g[illegible]; GSPDB:GN[illegible]; ATGP:A_TM02
C;Genetics:
A;Gene: ATSP:A_TM021B04.12
A;Map position: 5
A;Introns: 98/3; 40/3; 1[illegible]/3; 1[illegible]/2; 2[illegible]/3; 388/3
C;Superfamily: garden pea SAR DNA-binding protein

  Query Match 100.0%; Score 38; DB 2; Length 439;
  Best Local Similarity 66.7%; Pred. No. 1.9e+02;
  Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 WXXWHF 6
|::|||
Db 184 WFGWHF 189

Search completed: February 20, 2003, 10:50:02
Job time : 13 secs

GenCore version 5.1.3

OM protein - protein search, using sw model

Run on: February 20, 2003, 10:41:55 ; Search time 6.66667 Seconds
(without alignments)
37.329 Million cell updates/sec

Title: US-09-912-414-11
Perfect score: 38
Sequence: 1 WXXWHF 6

Scoring table: BLOSUM62DX
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	38	100.0	358	1	MTC1_CAUCR	Q45971 caulobacter
2	38	100.0	376	1	MTS1_RHIME	Q30569 rhizobium m
3	38	100.0	377	1	MTR1_BRUAB	Q30570 brucella ab
4	38	100.0	424	1	FD6C_SOYBN	P48628 glycine max
5	38	100.0	439	1	Y412_ARATH	O04658 arabidopsis
6	38	100.0	443	1	FD6C_BRANA	P48627 brassica na
7	38	100.0	448	1	FD6C_ARATH	P46312 arabidopsis
8	38	100.0	479	1	CATA_PSEPU	Q59714 pseudomonas
9	38	100.0	504	1	SIK1_YEAST	Q12460 saccharomyc
10	38	100.0	511	1	NOP5_YEAST	Q12499 saccharomyc
11	38	100.0	529	1	NOP5_HUMAN	Q9y2x3 homo sapien
12	38	100.0	534	1	NOP5_RAT	Q9qz86 rattus norv
13	38	100.0	540	1	Y42D_RHISN	P55386 rhizobium s
14	38	100.0	718	1	PLSB_CAEEL	Q20243 caenorhabdi
15	35	92.1	140	1	LYSA_DROME	P37157 drosophila
16	35	92.1	140	1	LYSB_DROME	Q29694 drosophila
17	35	92.1	140	1	LYSE_DROME	P37159 drosophila
18	35	92.1	140	1	LYSS_DROME	P37160 drosophila
19	35	92.1	142	1	LYSX_DROME	P37161 drosophila
20	35	92.1	372	1	YC60_MYCTU	Q11058 mycobacteri
21	35	92.1	396	1	BCR_ECOLI	P28246 escherichia
22	35	92.1	482	1	CATA_ONCVE	Q27712 onchocerca
23	35	92.1	483	1	5HT3_RAT	P35563 rattus norv
24	35	92.1	487	1	5HT3_MOUSE	P23979 mus musculu
25	33	86.8	170	1	LSPA_SERMA	Q52213 serratia ma
26	33	86.8	461	1	FUCO_HUMAN	P04066 homo sapien
27	33	86.8	462	1	FUCO_RAT	P17164 rattus norv
28	33	86.8	465	1	FUCO_CANFA	P48300 canis famil
29	33	86.8	475	1	YIFO_ECOLI	P31474 escherichia
30	33	86.8	1015	1	FDNG_ECOLI	P24183 escherichia
31	33	86.8	1016	1	FDOG_ECOLI	P32176 escherichia
32	33	86.8	1028	1	FDXG_HAEIN	P46448 haemophilus
33	33	86.8	3341	1	POLG_MCFA	P33515 m genome po
34	32	84.2	85	1	VG73_BPML5	Q05288 mycobacteri
35	32	84.2	103	1	ES6D_MYCTU	O05440 mycobacteri
36	32	84.2	105	1	NIGM_HUMAN	Q95179 homo sapien
37	32	84.2	108	1	NIGM_POVIN	Q02274 bos taurus
38	32	84.2	111	1	VPX_HV2KP	Q74122 human immun
39	32	84.2	111	1	YFBW_SALTI	P81891 salmonella
40	32	84.2	111	1	YFBW_SALTY	Q52328 salmonella
41	32	84.2	112	1	VPX_HV2CA	P24110 human immun
42	32	84.2	112	1	VPX_HV2D1	P17760 human immun
43	32	84.2	112	1	VPX_HV2G1	P18045 human immun
44	32	84.2	112	1	VPX_HV2N2	P05915 human immun
45	32	84.2	112	1	VPX_HV2RO	P06939 human immun

ALIGNMENTS

RESULT 1
MTC1_CAUCR
ID MTC1_CAUCR STANDARD; PRT; 358 AA.
AC Q45971;
DT 30-MAY-2000 (Rel. 39, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Modification methylase CcrmI (EC 2.1.1.72) (Adenine-specific
DE methyltransferase CcrmI) (M.CcrmI).
GN CCRMIM OR CCRM OR CC0378.
OS Caulobacter crescentus.
OC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;
OC Caulobacter.
OX NCBI_TaxID=155892;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CB15N / NA1000;
RX MEDLINE=94118303; PubMed=8289276;
RA Zweiger G., Marczynski G., Shapiro L.;
RT "A Caulobacter DNA methyltransferase that functions only in the
RT predivisional cell.";
RL J. Mol. Biol. 235:472-485(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 19089 / CB15;
RX MEDLINE=21173698; PubMed=11259647;
RA Nierman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
RA Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
RA Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
RA DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,
RA Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,
RA Utterback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,
RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
RT "Complete genome sequence of Caulobacter crescentus.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
CC -!- FUNCTION: THIS METHYLASE RECOGNIZES THE DOUBLE STRANDED SEQUENCE
CC GANTC AND CAUSES SPECIFIC METHYLATION ON A-2 ON BOTH STRANDS.
CC CCRM MEDIATED METHYLATION HAS IMPORTANT CELLULAR FUNCTIONS.
CC APPEARS TO CONTRIBUTE TO THE ACCURATE CELL CYCLE CONTROL OF DNA
CC REPLICATION AND CELLULAR MORPHOLOGY.
CC -!- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + DNA adenine = S-
CC adenosyl-L-homocysteine + DNA 6-methylaminopurine.
CC ----------
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC ----------
DR EMBL; U01032; AAA18913.1; -.
DR EMBL; AE005711; AAK22365.1; -.
DR HSSP; P11409; 1BOO.
DR REBASE; 2539; M.CcrMI.
DR TIGR; CC0378; -.

```
DR   InterPro; IPR001091; CN4_Metransf.
DR   InterPro; IPR002295; Cyt_C5_mtfrase.
DR   InterPro; IPR002941; N6/N4_Mtase.
DR   InterPro; IPR002052; N6_Mtase.
DR   InterPro; IPR000051; SAM_bind.
DR   Pfam; PF01555; N6_N4_Mtase; 1.
DR   PRINTS; PR00506; C5MTPRASE.
DR   PRINTS; PR00508; S21N4MTPRASE.
DR   PROSITE; PS00092; N6_MTASE; 1.
KW   Transferase; Methyltransferase; DNA replication; Complete proteome.
FT   CONFLICT    242    242       Y -> D (IN REF. 1).
SQ   SEQUENCE   376 AA,  42865 MW,  0DF4B20EF7D4D014 CRC64;

  Query Match             100.0%;  Score 38;  DB 1;  Length 376;
  Best Local Similarity   66.7%;  Prod. No. 43;
  Matches    4;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 WXXWHF 6
           |::|||
Db     330 WTYWHF 335


RESULT 2
MTS1_RHIME
ID   MTS1_RHIME     STANDARD;      PRT;   376 AA.
AC   O30569;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Modification methylase SmeIF (EC 2.1.1.72) (Adenine-specific
DE   methyltransferase SmeIF) (M.SmeI) (M.CcrMI).
GN   SMEIM OR CCRM OR R07C07.8 OR CMT00001.
OS   Rhizobium meliloti (Sinorhizobium meliloti).
OC   Bacteria, Proteobacteria, alpha subdivision, Rhizobiaceae group,
OC   Rhizobiaceae, Sinorhizobium.
OX   NCBI_TaxID=382,
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN 1021.
RX   MEDLINE=97440139; PubMed=9294447;
RA   Wright R., Stephens C., Shapiro L.;
RT   "The CcrM DNA methyltransferase is widespread in the alpha subdivision
RT   of proteobacteria, and its essential functions are conserved in
RT   Rhizobium meliloti and Caulobacter crescentus.";
RL   J. Bacteriol. 179:5869-5877(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=1021,
RX   MEDLINE=21394030; PubMed=11461148;
RA   Capela D., Barloy-Hubler F., Gouzy J., Bothe G., Ampe F., Batut J.,
RA   Boistard P., Becker A., Boutry M., Cadieu E., Dreano S., Gloux S.,
RA   Godrie T., Goffeau A., Kahn D., Kiss E., Lelaure V., Masuy D.,
RA   Pohl T., Portetelle D., Puhler A., Purnelle B., Ramsperger U.,
RA   Renard C., Thebault P., Vandenbol M., Weidner S., Galibert F.;
RT   "Analysis of the chromosome sequence of the legume symbiont
RT   Sinorhizobium meliloti strain 1021.",
RL   Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882(2001).
CC   -!- FUNCTION: THIS METHYLASE RECOGNIZES THE DOUBLE STRANDED SEQUENCE
CC       GANTC AND CAUSES SPECIFIC METHYLATION ON A C ON BOTH STRANDS.
CC       CCRM-MEDIATED METHYLATION HAS IMPORTANT CELLULAR FUNCTIONS.
CC       APPEARS TO CONTRIBUTE TO THE ACCURATE CELL CYCLE CONTROL OF DNA
CC       REPLICATION AND CELLULAR MORPHOLOGY.
CC   -!- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + DNA adenine = S-
CC       adenosyl-L-homocysteine + DNA 6-methylaminopurine.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AF011895; AAB71350.1; -.
DR   EMBL; AL591795; CAC45498.1; -.
DR   REBASE, 3264, M.SmeIF.
DR   InterPro, IPR001091, CN4_Metransf.
DR   InterPro; IPR002295, Cyt_C5_mtfrase.
DR   InterPro; IPR002941; N6/N4_Mtase.
DR   InterPro; IPR002052, N6_Mtase.
DR   InterPro; IPR000051; SAM_bind.
DR   Pfam; PF01555; N6_N4_Mtase; 1.
DR   PRINTS; PR00506; C5MTPRASE.
DR   PRINTS; PR00508; S21N4MTPRASE.
DR   PROSITE; PS00092; N6_MTASE; 1.
KW   Transferase; Methyltransferase; DNA replication; Complete proteome.
FT   CONFLICT    125    141       MPMFNPK -> IPAELLQ (IN REF. 1).
FT   CONFLICT    157    157       P -> A (IN REF. 1).
SQ   SEQUENCE   376 AA,  42865 MW,  0DF4B20EF7D4D014 CRC64;

  Query Match             100.0%;  Score 38;  DB 1;  Length 376;
  Best Local Similarity   66.7%;  Prod. No. 52;
  Matches    4;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 WXXWHF 6
           |::|||
Db     347 WTFWHF 352


RESULT 3
MTB1_BRUAB
ID   MTB1_BRUAB     STANDARD;      PRT;   377 AA.
AC   O30070;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Modification methylase BabI (EC 2.1.1.72) (Adenine-specific
DE   methyltransferase BabI) (M.BabI) (M.CcrMI).
GN   BABIM OR CCRM.
OS   Brucella abortus.
OC   Bacteria, Proteobacteria, alpha subdivision, Rhizobiaceae group;
OC   Brucellaceae, Brucella.
OX   NCBI_TaxID=235,
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S2308;
RX   MEDLINE=97440139; PubMed=9294447;
RA   Wright R., Stephens C., Shapiro L.;
RT   "The CcrM DNA methyltransferase is widespread in the alpha subdivision
RT   of proteobacteria, and its essential functions are conserved in
RT   Rhizobium meliloti and Caulobacter crescentus.";
RL   J. Bacteriol. 179:5869-5877(1997).
CC   -!- FUNCTION: THIS METHYLASE RECOGNIZES THE DOUBLE STRANDED SEQUENCE
CC       GANTC AND CAUSES SPECIFIC METHYLATION ON A C ON BOTH STRANDS.
CC       CCRM-MEDIATED METHYLATION HAS IMPORTANT CELLULAR FUNCTIONS.
CC       APPEARS TO CONTRIBUTE TO THE ACCURATE CELL-CYCLE CONTROL OF DNA
CC       REPLICATION AND CELLULAR MORPHOLOGY.
CC   -!- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + DNA adenine = S-
CC       adenosyl-L-homocysteine + DNA 6-methylaminopurine.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AF011895; AAB71351.1; -.
DR   PREBASE; 3263; M.BabI
DR   InterPro, IPR001091, CN4_Metransf.
DR   InterPro, IPR002295, Cyt_C5_mtfrase.
DR   InterPro; IPR002941; N6/N4_Mtase.
DR   InterPro, IPR002052, N6_Mtase.
DR   InterPro; IPR000051; SAM_bind.
DR   Pfam; PF01555; N6_N4_Mtase; 1.
```

```
DR   PRINTS; PR00506; D21N6MTFRASE.
DR   PRINTS; PR00508; S21N4MTFRASE.
DR   PROSITE; PS00092; N6_MTASE; 1.
KW   Transferase; Methyltransferase; DNA replication.
SQ   SEQUENCE   377 AA;  42202 MW;  697C88A25580B30D CRC64;

  Query Match           100.0%;  Score 38;  DB 1;  Length 377;
  Best Local Similarity   66.7%;  Pred. No. 52;
  Matches    4;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 WXXWHF 6
          |::|||
Db    347 WTFWHF 352


RESULT 4
FD6C_SOYBN
ID   FD6C_SOYBN     STANDARD;      PRT;   424 AA.
AC   P48628;
DT   01-FEB-1996 (Rel. 33, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Omega-6 fatty acid desaturase, chloroplast precursor (EC 1.14.99.-).
OS   Glycine max (Soybean)
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine
OX   NCBI_TaxID=3847;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Seed;
RX   MEDLINE=94345008; PubMed=8066133;
RA   Hitz W.D., Carlson T.J., Booth J.R. Jr., Kinney A.J., Stecca K.L.,
RA   Yadav N.S.;
RT   "Cloning of a higher-plant plastid omega-6 fatty acid desaturase cDNA
RT   and its expression in a cyanobacterium.";
RL   Plant Physiol. 105:635-641(1994).
CC   -!- FUNCTION: CHLOROPLAST OMEGA-6 FATTY ACID DESATURASE INTRODUCES
CC       THE SECOND DOUBLE BOND IN THE BIOSYNTHESIS OF 18:3 AND 18:3 FATTY
CC       ACIDS. IMPORTANT CONSTITUENTS OF PLANT MEMBRANES. IT IS THOUGHT
CC       TO USE FERREDOXIN AS AN ELECTRON DONOR AND TO ACT ON FATTY ACIDS
CC       ESTERIFIED TO GALACTOLIPIDS, SULFOLIPIDS AND PHOSPHATIDYLGLYCEROL.
CC   -!- PATHWAY: Polyunsaturated fatty acid biosynthesis.
CC   -!- SUBCELLULAR LOCATION: CHLOROPLAST, MEMBRANE BOUND (PROBABLE).
CC   -!- DOMAIN: THE HISTIDINE BOX DOMAINS MAY CONTAIN THE ACTIVE SITE
CC       AND/OR BE INVOLVED IN METAL ION BINDING.
CC   -!- SIMILARITY: BELONGS TO THE FATTY ACID DESATURASE FAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; L29215; AAA50189.1; -.
DR   InterPro; IPR001225; FA_desaturase.
DR   Pfam; PF00487; FA_desaturase; 1.
DR   ProDom; PD001081; FA_desaturase; 1.
KW   Oxidoreductase; Fatty acid biosynthesis; Chloroplast; Membrane;
KW   Transit peptide.
FT   TRANSIT       1     63       CHLOROPLAST (BY SIMILARITY).
FT   CHAIN        64    424       OMEGA-6 FATTY ACID DESATURASE.
FT   DOMAIN      165    169       HISTIDINE BOX-1.
FT   DOMAIN      201    205       HISTIDINE BOX-2.
FT   DOMAIN      361    365       HISTIDINE BOX-3.
SQ   SEQUENCE   424 AA;  49641 MW;  4F2DBF2D4B1A2000 CRC64;

  Query Match           100.0%;  Score 38;  DB 1;  Length 424;
  Best Local Similarity   66.7%;  Pred. No. 58;
  Matches    4;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 WXXWHF 6
          |::|||
Db    251 WLMWHF 256


RESULT 5
Y412_ARATH
ID   Y412_ARATH     STANDARD;      PRT;   439 AA.
AC   O04658;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Hypothetical protein At5g27120.
GN   AT5G27120 OR TMO21B04.12 OR T21B4.30.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Columbia;
RX   MEDLINE=21016721; PubMed=11130714;
RA   Tabata S., Kaneko T., Nakamura Y., Kotani H., Kato T., Asamizu E.,
RA   Miyajima N., Sasamoto S., Kimura T., Hosouchi T., Kawashima K.,
RA   Kohara M., Matsumoto M., Matsuno A., Muraki A., Nakayama S.,
RA   Nakazaki N., Naruo K., Okumura S., Shinpo S., Takeuchi C., Wada T.,
RA   Watanabe A., Yamada M., Yasuda M., Sato S., de la Bastide M.,
RA   Huang E., Spiegel L., Gnoj L., O'Shaughnessy A., Preston R.,
RA   Habermann K., Murray J., Johnson D., Rohlfing T., Nelson J.,
RA   Stoneking T., Pepin K., Spieth J., Sekhon M., Armstrong J., Becker M.,
RA   Belter E., Cordum H., Cordes M., Courtney L., Courtney W., Dante M.,
RA   Du H., Edwards J., Fryman J., Haakensen B., Lamar E., Latreille P.,
RA   Leonard S., Meyer R., Mulvaney E., Ozersky P., Riley A., Strowmatt C.,
RA   Wagner-McPherson C., Wollam A., Yoakum M., Bell M., Dedhia N.,
RA   Parnell L., Shah R., Rodriguez M., Hoon See L., Vil D., Baker J.,
RA   Kirchoff K., Toth K., King L., Bahret A., Miller B., Marra M.,
RA   Martienssen R., McCombie W.R., Wilson R.K., Murphy G., Bancroft I.,
RA   Volckaert G., Wambutt R., Duesterhoeft A., Stiekema W., Pohl T.,
RA   Entian K.-D., Terryn N., Hartley N., Bent E., Johnson S.,
RA   Langham S.-A., McCullagh B., Robben J., Grymonprez B., Zimmermann W.,
RA   Ramsperger U., Wedler H., Balke K., Wedler E., Peters S.,
RA   van Staveren M., Dirkse W., Mooijman P., Klein Lankhorst R.,
RA   Weitzenegger T., Bothe G., Rose M., Hauf J., Berneiser S., Hempel S.,
RA   Feldpausch M., Lamberth S., Villarroel R., Gielen J., Ardiles W.,
RA   Bents O., Lemcke K., Kolesov G., Mayer K., Rudd S., Schoof H.,
RA   Schueller C., Zaccaria P., Mewes H.-W., Bevan M., Fransz P.;
RT   "Sequence and analysis of chromosome 5 of the plant Arabidopsis
RT   thaliana.";
RL   Nature 408:823-826(2000).
CC   -!- SIMILARITY: BELONGS TO THE NOP5/NOP56 FAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AF007271; AAB61073.1; -.
DR   InterPro; IPR002687; Nop.
DR   Pfam; PF01798; Nop; 1.
DR   ProDom; PD004104; Nop; 1.
KW   Hypothetical protein.
SQ   SEQUENCE   439 AA;  47932 MW;  BE0E2214F9EC4FD8 CRC64;

  Query Match           100.0%;  Score 38;  DB 1;  Length 439;
  Best Local Similarity   66.7%;  Pred. No. 59;
  Matches    4;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 WXXWHF 6
          |::|||
```

```
Db    194 WFGWHF 193


RESULT 6
FEL2_BRANA
ID   FD6C_BRANA     STANDARD;      PRT;   443 AA.
AC   P48627;
DT   01-FEB-1996 (Rel. 33, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Omega-6 fatty acid desaturase, chloroplast precursor (EC 1.14.99.-).
OS   Brassica napus (Rape).
OC   Eukaryota; Viridiplantae, Streptophyta, Embryophyta, Tracheophyta,
OC   Spermatophyta, Magnoliophyta, eudicotyledons, core eudicots, Rosidae,
OC   eurosids II, Brassicales, Brassicaceae; Brassica.
OX   NCBI_TaxID=3708;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Seed;
RX   MEDLINE=94134008; PubMed=8260133;
RA   Hitz W.D., Carlson T.J., B[illegible] F., Kinney A.J., Stecca K.L.,
RA   Yadav N.S.;
RT   "Cloning of a higher-plant plastid omega-6 fatty acid desaturase cDNA
RT   and its expression in a cyanobacterium.";
RL   Plant Physiol. 105:635-641(1994).
CC   -!- FUNCTION: CHLOROPLAST OMEGA-6 FATTY ACID DESATURASE INTRODUCES
CC       THE SECOND DOUBLE BOND IN THE BIOSYNTHESIS OF 16:3 AND 18:3 FATTY
CC       ACIDS, IMPORTANT CONSTITUENTS OF PLANT MEMBRANES. IT IS THOUGHT
CC       TO USE FERREDOXIN AS AN ELECTRON DONOR AND TO ACT ON FATTY ACIDS
CC       ESTERIFIED TO GALACTOLIPIDS, SULFOLIPIDS AND PHOSPHATIDYLGLYCEROL.
CC   -!- PATHWAY: Polyunsaturated fatty acid biosynthesis.
CC   -!- SUBCELLULAR LOCATION: Membrane-bound (Probable).
CC   -!- DOMAIN: THE HISTIDINE BOX DOMAINS MAY CONTAIN THE ACTIVE SITE
CC       AND/OR BE INVOLVED IN METAL ION BINDING.
CC   -!- SIMILARITY: BELONGS TO THE FATTY ACID DESATURASE FAMILY.
CC   -----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; L08214; AAA33197.1; -.
DR   InterPro; IPR001199; FA_desaturase.
DR   Pfam; PF00487; FA_desaturase; 1.
DR   ProDom; PD001081; FA_desaturase; 1.
KW   Oxidoreductase; Fatty acid biosynthesis; Chloroplast; Membrane;
KW   Transit peptide.
FT   TRANSIT        1     64       CHLOROPLAST (BY SIMILARITY).
FT   CHAIN         65    443       OMEGA-6 FATTY ACID DESATURASE.
FT   DOMAIN       127    131       HISTIDINE BOX 1.
FT   DOMAIN       [illegible]       HISTIDINE BOX 2.
FT   DOMAIN       362    366       HISTIDINE BOX 3.
SQ   SEQUENCE   443 AA;  [illegible] MW;  [illegible];

  Query Match             [illegible]; Score 36;  DB 1;  Length 443;
  Best Local Similarity;   66.7%;  Pred. No. 61;
  Matches    4;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 WXXWHF 6
           :::|:|
Db     252 WVNWHF 257


RESULT 7
FD6C_ARATH
ID   FD6C_ARATH     STANDARD;      PRT;   448 AA.
AC   P46312; [illegible];
DT   01-NOV-1995 (Rel. 32, Created)
DT   16-OCT-2001 (Rel. 41, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Omega-6 fatty acid desaturase, chloroplast precursor (EC 1.14.99.-).
GN   FAD6 OR FAD6 OR AT4G30950 OR F6I18.140.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae, Streptophyta, Embryophyta, Tracheophyta;
OC   Spermatophyta, Magnoliophyta; eudicotyledons, core eudicots, Rosidae;
OC   eurosids II, Brassicales, Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE OF 1-418 FROM N.A.
RC   STRAIN=cv. Columbia;
RX   MEDLINE=96146736; PubMed=7846158;
RA   Falcone D.L., Gibson S., Lemieux B., Somerville C.R.;
RT   "Identification of a gene that complements an Arabidopsis mutant
RT   deficient in chloroplast omega-3 desaturase activity.";
RL   Plant Physiol. 106:1453-1459(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Columbia;
RX   MEDLINE=20088488; PubMed=10617198;
RA   Mayer K.F.X., Schueller C., Wambutt R., Murphy G., Volckaert G.,
RA   Pohl T., Duesterhoeft A., Stiekema W., Entian K.-D., Terryn N.,
RA   Harris B., Ansorge W., Brandt P., Grivell L., Rieger M.,
RA   Weichselgartner M., de Simone V., Obermaier B., Mache R., Mueller M.,
RA   Kreis M., Delseny M., Puigdomenech P., Watson M., Schmidtheini T.,
RA   Reichert B., Portetelle D., Perez-Alonso M., Boutry M., Bancroft I.,
RA   Vos P., Hoheisel J., Zimmermann W., Wedler H., Ridley P.,
RA   Langham S.-A., McCullagh B., Bilham L., Robben J.,
RA   Van der Schueren J., Grymonprez B., Chuang Y.-J., Vandenbussche F.,
RA   Braeken M., Weltjens I., Voet M., Bastiaens I., Aert R., Defoor E.,
RA   Weitzenegger T., Bothe G., Ramsperger U., Hilbert H., Braun M.,
RA   Holzer E., Brandt A., Peters C., van Staveren M., Dirkse W.,
RA   Mooijman P., Klein Lankhorst R., Rose M., Hauf J., Koetter P.,
RA   Berneiser S., Hempel S., Feldpausch M., Lamberth S., Van den Daele H.,
RA   De Keyser A., Buysshaert C., Gielen J., Villarroel R., De Clercq R.,
RA   Van Montagu M., Rogers J., Cronin A., Quail M., Bray-Allen S.,
RA   Clark L., Doggett J., Hall S., Kay M., Lennard N., McLay K., Mayes R.,
RA   Pettett A., Rajandream M.A., Lyne M., Benes V., Rechmann S.,
RA   Borkova D., Blocker H., Scharfe M., Grimm M., Loehnert T.-H.,
RA   Dose S., de Haan M., Maarse A., Schaefer M., Mueller-Auer S.,
RA   Gabel C., Fuchs M., Fartmann B., Granderath K., Dauner D., Herzl A.,
RA   Neumann S., Argiriou A., Vitale D., Liguori R., Piravandi E.,
RA   Massenet O., Quigley F., Clabauld G., Muendlein A., Felber R.,
RA   Schnabl S., Hiller R., Schmidt W., Lecharny A., Aubourg S.,
RA   Chefdor F., Cooke R., Berger C., Montfort A., Casacuberta E.,
RA   Gibbons T., Weber N., Vandenbol M., Bargues M., Terol J., Torres A.,
RA   Perez-Perez A., Purnelle B., Bent E., Johnson S., Tacon D., Jesse T.,
RA   Heijnen L., Schwarz S., Scholler P., Heber S., Francs P., Bielke C.,
RA   Frishman D., Haase D., Lemcke K., Mewes H.-W., Stocker S.,
RA   Zaccaria P., Bevan M., Wilson R.K., de la Bastide M., Habermann K.,
RA   Parnell L., Dedhia N., Gnoj L., Schutz K., Huang E., Spiegel L.,
RA   Sekhon M., Murray J., Sheet P., Cordes M., Abu-Threideh J.,
RA   Stoneking T., Kalicki J., Graves T., Harmon G., Edwards J.,
RA   Latreille P., Courtney L., Cloud J., Abbott A., Scott K., Johnson D.,
RA   Minx P., Bentley D., Fulton B., Miller N., Greco T., Kemp K.,
RA   Kramer J., Fulton L., Mardis E., Dante M., Pepin K., Hillier L.,
RA   Nelson J., Spieth J., Ryan E., Andrews S., Geisel C., Layman D.,
RA   Du H., Ali J., Berghoff A., Jones K., Drone K., Cotton M., Joshu C.,
RA   Antonoiu B., Zidanic M., Strong C., Sun H., Lamar B., Yordan C.,
RA   Ma P., Zhong J., Preston R., Vil D., Shekher M., Matero A., Shah R.,
RA   Swaby I.K., O'Shaughnessy A., Rodriguez M., Hoffman J., Till S.,
RA   Granat S., Shohdy N., Hasegawa A., Hameed A., Lodhi M., Johnson A.,
RA   Chen E., Marra M., Martienssen R., McCombie W.R.;
RT   "Sequence and analysis of chromosome 4 of the plant Arabidopsis
RT   thaliana.";
RL   Nature 402:769-777(1999).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Columbia;
RA   Shinozaki K., Davis R.W., Federspiel N.A., Theologis A.,
RT   "RIKEN Arabidopsis full-length cDNA clones (RAFL) sequenced by the
RT   SSP consortium (Salk/Stanford/PGEC).";
RL   Submitted (SEP-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: CHLOROPLAST OMEGA-6 FATTY ACID DESATURASE INTRODUCES
```

```
CC       THE SECOND DOUBLE BOND IN THE BIOSYNTHESIS OF 16.3 AND 18.3 FATTY
CC       ACIDS, IMPORTANT CONSTITUENTS OF PLANT MEMBRANES. IT IS THOUGHT
CC       TO USE FERREDOXIN AS AN ELECTRON DONOR AND TO ACT ON FATTY ACIDS
CC       ESTERIFIED TO GALACTOLIPIDS, SULFOLIPIDS AND PHOSPHATIDYLGLYCEROL.
CC   -!- PATHWAY: Polyunsaturated fatty acid biosynthesis.
CC   -!- SUBCELLULAR LOCATION: CHLOROPLAST, MEMBRANE-BOUND (PROBABLE).
CC   -!- DEVELOPMENTAL STAGE: HIGHEST LEVELS FOUND IN EXPANDING LEAVES.
CC   -!- DOMAIN: THE HISTIDINE BOX DOMAINS MAY CONTAIN THE ACTIVE SITE
CC       AND/ OR BE INVOLVED IN METAL ION BINDING.
CC   -!- SIMILARITY: BELONGS TO THE FATTY ACID DESATURASE FAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaborati
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; U09503; AAA92800.1; -.
DR   EMBL; AL022198; CAA18198.1; -.
DR   EMBL; AL161578; CAB79813.1; -.
DR   EMBL; AY045621; AAK73979.1; -.
DR   EMBL; AY058078; AAL24186.1; -.
DR   EMBL; AY058852; AAL24240.1; -.
DR   InterPro; IPR001225; FA_desaturase.
DR   Pfam; PF00487; FA_desaturase; 1.
DR   ProDom; PD001081; FA_desaturase; 1.
KW   Oxidoreductase; Fatty acid biosynthesis; Chloroplast; Membrane;
KW   Transit peptide.
FT   TRANSIT       1     69       CHLOROPLAST (BY SIMILARITY).
FT   CHAIN        70    448       OMEGA-6 FATTY ACID DESATURASE.
FT   DOMAIN      171    175       HISTIDINE BOX-1.
FT   DOMAIN      207    211       HISTIDINE BOX-2.
FT   DOMAIN      367    371       HISTIDINE BOX-3.
SQ   SEQUENCE   448 AA;  51225 MW;  C2AC72FB29FBF287 CRC64;

  Query Match             100.0%;  Score 38;  DB 1;  Length 448;
  Best Local Similarity   66.7%;  Pred. No. 61;
  Matches    4;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 WXXWHF 6
           |::|||
Db     257 WVNWHF 262


RESULT 8
CATA_PSEPU
ID   CATA_PSEPU     STANDARD;      PRT;   479 AA.
AC   Q59714;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   Catalase (EC 1.11.1.6).
GN   KATA OR CATA.
OS   Pseudomonas putida
OC   Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC   Pseudomonas.
OX   NCBI_TaxID=303;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Corvallis;
RX   MEDLINE=98019091; PubMed=9358059;
RA   Kim Y.C., Miller C.D., Anderson A.J.;
RT   "Identification of adjacent genes encoding the major catalase and a
RT   bacterioferritin from the plant-beneficial bacterium Pseudomonas
RT   putida.";
RL   Gene 199:219-224(1997).
CC   -!- FUNCTION: DECOMPOSES HYDROGEN PEROXIDE IN WATER AND OXYGEN, SERVES
CC       TO PROTECT CELLS FROM THE TOXIC EFFECTS OF HYDROGEN PEROXIDE.
CC   -!- CATALYTIC ACTIVITY: 2 H(2)O(2) = O(2) + 2 H(2)O.
CC   -!- COFACTOR: HEME GROUP.
CC   -!- ENZYME REGULATION: ACTIVATED BY PEROXIDE.
CC   -!- SIMILARITY: BELONGS TO THE CATALASE FAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaborati
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; U63511; AAB88219.1; -.
DR   HSSP; P42321; 2CAE.
DR   InterPro; IPR002226; Catalase.
DR   Pfam; PF00199; catalase; 1.
DR   PRINTS; PR00067; CATALASE.
DR   ProDom; PD000510; Catalase; 1.
DR   PROSITE; PS00437; CATALASE_1; 1.
DR   PROSITE; PS00438; CATALASE_2; 1.
KW   Oxidoreductase; Peroxidase; Iron; Heme; Hydrogen peroxide.
FT   ACT_SITE     53     53       BY SIMILARITY.
FT   ACT_SITE    126    126       BY SIMILARITY.
FT   BINDING     336    336       PROXIMAL HEME LIGAND (BY SIMILARITY).
SQ   SEQUENCE   479 AA;  53381 MW;  EFE3CBDE67778571 CRC64;

  Query Match             100.0%;  Score 38;  DB 1;  Length 479;
  Best Local Similarity   66.7%;  Pred. No. 64;
  Matches    4;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 WXXWHF 6
           |::|||
Db     209 WVKWHF 214


RESULT 9
SIK1_YEAST
ID   SIK1_YEAST     STANDARD;      PRT;   504 AA.
AC   Q12460;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   SIK1 protein (Nucleolar protein NOP56).
GN   SIK1 OR NOP56 OR YLR197W OR L8167.9.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288c / YPH1;
RX   MEDLINE=96040178; PubMed=7547500;
RA   Morin P.J., Downs J.A., Snodgrass A.M., Gilmore T.D.;
RT   "Genetic analysis of growth inhibition by GAL4-I kappa B-alpha in
RT   Saccharomyces cerevisiae.";
RL   Cell Growth Differ. 6:789-798(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288c / AB972;
RA   Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Du Z.,
RA   Favello A., Fulton L., Gattung S., Greco T., Kirsten J., Korabia T.,
RA   Hallsworth K., Hawkins J., Hillier L., Jier M., Johnson D.,
RA   Johnston L., Langston Y., Latreille P., Mardis E., Menezes S.,
RA   Miller N., Nhan M., Pauley A., Peluso D., Rifken L., Riles L.,
RA   Taich S., Trevaskis E., Vaudin M., Vignati D., Wilcox L., Wilson R.,
RA   Wohldman P., Waterston R.;
RL   Submitted (SEP-1994) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   CHARACTERIZATION, AND MUTAGENESIS.
RX   MEDLINE=98038777; PubMed=9372940;
RA   Gautier T., Berges T., Tollervey D., Hurt E.;
RT   "Nucleolar KKE/D repeat proteins Nop56p and Nop58p interact with Nop1p
RT   and are required for ribosome biogenesis.";
RL   Mol. Cell. Biol. 17:7088-7098(1997).
CC   -!- FUNCTION: REQUIRED FOR 60S RIBOSOMAL SUBUNIT SYNTHESIS.
```

```
CC   -!- SUBUNIT: INTERACTS WITH NOP1 AND NOP56.
CC   -!- SUBCELLULAR LOCATION: Nuclear, nucleolar.
CC   -!- SIMILARITY: BELONGS TO THE NOP5/NOP58 FAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; U12277; AAC49066.1; -.
DR   EMBL; U14913; AAB67451.1; -.
DR   SGD; S0...4167; SIK1.
DR   InterPro; IPR002687; Nop.
DR   Pfam; PF01798; Nop; 1.
DR   ProDom; PD004104; Nop; 1.
KW   Ribosome biogenesis, Nuclear protein.
FT   DOMAIN       443     504       ASP/GLU/LYS-RICH.
FT   MUTAGEN      333     333       V->A: REDUCED GROWTH RATE AT ALL
FT                                  TEMPERATURES, WHEN ASSOCIATED WITH R-385.
FT   MUTAGEN      385     385       Y->C: AT 37 DEGREES, GROWTH SLOWS AFTER 6
FT                                  TO 8 HOURS AND CELL DIVISION STOPS AFTER
FT                                  20 HOURS.
FT   MUTAGEN      385     385       M->R: REDUCED GROWTH RATE AT ALL
FT                                  TEMPERATURES, WHEN ASSOCIATED WITH A 333.
SQ   SEQUENCE   504 AA,  58864 MW,  FE2E2A5970B14842 CRC64;

  Query Match             100.0%;  Score 38;  DB 1;  Length 504;
  Best Local Similarity   66.7%;  Pred. No. 6;
  Matches    4;  Conservative    2;  Mismatches    1;  Indels    1;  Gaps    1;

Qy       1 WXXWHF 6
           |::|||
Db     190 WYGWHF 203


RESULT 10
NOP5_YEAST
ID   NOP5_YEAST     STANDARD;      PRT;   511 AA.
AC   P21759;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Nucleolar protein NOP5 (Nucleolar protein 5) (NOP58).
GN   NOP5 OR NOP58 OR YOR310C OR O9120.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycetales; Saccharomycetaceae;
OC   Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288c / FY1679;
RA   Pearson B.M., Hermand Y., Wolf S.S., Baldacci-Cresp F., Lecluse A.,
RA   Simon M.-N. (AUG-1997) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A., AND CHARACTERIZATION.
RC   STRAIN=W303;
RX   MEDLINE=98098105; PubMed=9402712;
RA   Wu P., Brockenbrough J.S., Metcalfe A.C., Chen S., Aris J.P.;
RT   "Nop5p is a small nucleolar ribonucleoprotein component required for
RT   pre-18S rRNA processing in yeast.";
RL   J. Biol. Chem. 273:16453-16460(1998).
CC   -!- FUNCTION: REQUIRED FOR PRE-18S RRNA PROCESSING. MAY BIND
CC       MICROTUBULES.
CC   -!- SUBUNIT: INTERACTS WITH NOP56 AND NOP1.
CC   -!- SUBCELLULAR LOCATION: Nuclear; nucleolar.
CC   -!- SIMILARITY: BELONGS TO THE NOP5/NOP58 FAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaborat
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; X99565; CAA67165.1; -.
DR   EMBL; Z75217; CAA99630.1; -.
DR   EMBL; AF056979; AAC39484.1; -.
DR   SGD; S0005937; NOP58.
DR   InterPro; IPR002687; Nop.
DR   Pfam; PF01798; Nop; 1.
DR   ProDom; PD004104; Nop; 1.
KW   Ribosome biogenesis, Nuclear protein, tRNA processing.
FT   DOMAIN       441     511       ASP/GLU/LYS-RICH.
SQ   SEQUENCE   511 AA,  [illegible] MW;  [illegible] CRC64;

  Query Match             [illegible]%;  Score 38;  DB 1;  Length 511;
  Best Local Similarity   66.7%;  Pred. No. 6;
  Matches    4;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 WXXWHF 6
           |::|||
Db     187 WYGWHF 192


RESULT 11
NOP5_HUMAN
ID   NOP5_HUMAN     STANDARD;      PRT;   529 AA.
AC   Q9Y2X3; Q9P036; Q9UFN3;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Nucleolar protein NOP5 (Nucleolar protein 5) (NOP58) (HSPC120).
GN   NOP5.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Lyman S.K., Gerace L.;
RT   "Cloning and characterization of NOP5/NOP58.";
RL   Submitted (JAN-1999) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20394617; PubMed=10925035;
RA   Nielsen S.A., Sandberg F.F., LaRochelle W.J.;
RT   "Isolation and characterization of a novel PDGF-induced human gene.";
RL   Gene [illegible] (2000).
RN   [3]
RP   SEQUENCE OF 1-447 FROM N.A.
RC   TISSUE=Brain;
RA   Bloecker H., Boecher M., Brandt P., Mewes H.-W., Gassenhuber J.,
RA   Wiemann S.;
RL   Submitted (APR-1999) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE OF 64-529 FROM N.A.
RC   TISSUE=Blood;
RX   MEDLINE=[illegible]; PubMed=11042152;
RA   Zhang Q. H., Ye M., Wu X. Y., Ren S. X., Zhao M., Zhao C. J., Fu G.,
RA   Shen Y., Fan H. Y., Lu G., Zhong M., Xu X. R., Han Z. G., Zhang J.-W.,
RA   Tao J., Huang Q. H., Zhou J., Hu G. X., Gu J., Chen S.-J., Chen Z.;
RT   "Cloning and functional analysis of cDNAs with open reading frames for
RT   300 previously undefined genes expressed in CD34+ hematopoietic
RT   stem/progenitor cells.";
RL   Genome Res. 10:1546-1560(2000).
CC   -!- FUNCTION: REQUIRED FOR PRE-18S RIBOSOMAL RNA/SUBUNIT BIOGENESIS (By
CC       similarity).
CC   -!- SUBCELLULAR LOCATION: Nuclear, nucleolar.
CC   -!- TISSUE SPECIFICITY: Ubiquitous.
CC   -!- SIMILARITY: BELONGS TO THE NOP5/NOP58 FAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
```

```
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ------------------------------------------------------------------------
DR   EMBL; AF122534; AAD27610.1; -.
DR   EMBL; AF263608; AAF91394.1; -.
DR   EMBL; AL117554; CAB55989.1; -.
DR   EMBL; AF161469; AAF29094.1; -.
DR   SWISS-2DPAGE; Q9Y2X3; HUMAN.
DR   InterPro; IPR002687; Nop.
DR   Pfam; PF01798; Nop; 1.
DR   ProDom; PD004104; Nop; 1.
KW   Ribosome biogenesis; Nuclear protein.
FT   CONFLICT      2      2       L -> M (IN REF. 3).
FT   CONFLICT    129    129       G -> V (IN REF. 3).
FT   CONFLICT    ...    ...       LTYCFCLQFVGDFFNYASAF -> YHTASVYFFLAIGRLCL
FT                                CQ (IN REF. 4).
FT   CONFLICT    235    260       KAAAEISMGTEVSEEDICNILHLCTQ -> EGSCRDIHGNP
FT                                GFRRRYLQYSASLHP (IN REF. 4).
FT   CONFLICT    280    280       M -> V (IN REF. 3).
FT   CONFLICT    443    447       RKIEQ -> TQNTT (IN REF. 3).
SQ   SEQUENCE   529 AA;  59579 MW;  27CD23CFF5B0A556 CRC64;

  Query Match           100.0%;  Score 38;  DB 1;  Length 529;
  Best Local Similarity   66.7%;  Pred. No. 70;
  Matches    4;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 WXXWHF 6
           |::|||
Db     186 WYGWHF 191


RESULT 12
NOP5_RAT
ID   NOP5_RAT       STANDARD;      PRT;   534 AA.
AC   Q9QZ86; Q88525;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Nucleolar protein NOP5 (Nucleolar protein 5) (Nopp140-associated
DE   protein).
GN   NOP5 OR NAP65.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A., SEQUENCE OF 1-22, AND INTERACTION WITH NOLC1
RX   MEDLINE=20143579; PubMed=10679015;
RA   Yang Y., Isaac C., Wang C., Dragon F., Pogacic V., Meier U.T.;
RT   "Conserved composition of mammalian box H/ACA and box C/D small
RT   nucleolar ribonucleoprotein particles and their interaction with the
RT   common factor Nopp140.";
RL   Mol. Biol. Cell 11:567-577(2000).
RN   [2]
RP   SEQUENCE OF 1-461 FROM N.A.
RA   Hatton D., Gray J.C.;
RL   Submitted (JUN-1998) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: REQUIRED FOR 60S RIBOSOMAL SUBUNIT BIOGENESIS (By
CC       similarity).
CC   -!- SUBUNIT: Interacts with Nolc1/Nopp140.
CC   -!- SUBCELLULAR LOCATION: Nuclear, nucleolar.
CC   -!- SIMILARITY: BELONGS TO THE NOP5/NOP56 FAMILY.
CC   ------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ------------------------------------------------------------------------
DR   EMBL; AF134371; AAF05769.1; -.
DR   EMBL; AF069782; AAC23535.1; -.
DR   InterPro; IPR002687; Nop.
DR   Pfam; PF01798; Nop; 1.
DR   ProDom; PD004104; Nop; 1.
KW   Ribosome biogenesis; Nuclear protein.
FT   CONFLICT    396    396       R -> K (IN REF. 2).
FT   CONFLICT    459    459       A -> K (IN REF. 2).
SQ   SEQUENCE   534 AA;  60070 MW;  4B9585FA14E67739 CRC64;

  Query Match           100.0%;  Score 38;  DB 1;  Length 534;
  Best Local Similarity   66.7%;  Pred. No. 71;
  Matches    4;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 WXXWHF 6
           |::|||
Db     186 WYGWHF 191


RESULT 13
Y4CD_RHISN
ID   Y4CD_RHISN     STANDARD;      PRT;   640 AA.
AC   P55386;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   Hypothetical 71.6 kDa protein Y4CD.
GN   Y4CD.
OS   Rhizobium sp. (strain NGR234).
OG   Plasmid sym pNGR234a.
OC   Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC   Rhizobiaceae; Rhizobium
OX   NCBI_TaxID=394;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97305956; PubMed=9163424;
RA   Freiberg C.A., Pellay R., Bairoch A., Broughton W.J., Rosenthal A.,
RA   Perret X.;
RT   "Molecular basis of symbiosis between Rhizobium and legumes.";
RL   Nature 387:394-401(1997).
CC   -!- SIMILARITY: NONE OBVIOUS.
CC   ------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ------------------------------------------------------------------------
DR   EMBL; AE000067; AAB91634.1; -.
DR   InterPro; IPR000157; TIR_domain.
DR   SMART; SM00255; TIR; 1.
KW   Hypothetical protein; Plasmid.
SQ   SEQUENCE   640 AA;  71609 MW;  8[illegible] CRC64;

  Query Match           100.0%;  Score 38;  DB 1;  Length 640;
  Best Local Similarity   66.7%;  Pred. No. 84;
  Matches    4;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 WXXWHF 6
           |::|||
Db     521 WPFWHF 526


RESULT 14
PLSB_CAEEL
ID   PLSB_CAEEL     STANDARD;      PRT;   718 AA.
AC   Q22949;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
```

```
DT   15-JUL-1999 (Rel. 39, Last annotation update)
DE   Probable glycerol-3-phosphate acyltransferase, mitochondrial precursor
DE   (EC 2.3.1.15) (GPAT).
GN   gpat-2.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Bristol N2;
RA   Blumenthal T., Bradshaw H.;
RL   Submitted (JUL-1999) to the EMBL/GenBank/DDBJ databases.
CC   -!- CATALYTIC ACTIVITY: Acyl-CoA + sn-glycerol 3-phosphate = CoA +
CC       acyl-sn-glycerol 3-phosphate.
CC   -!- PATHWAY: FIRST STEP IN DE NOVO PHOSPHOLIPID BIOSYNTHESIS. IT MAY
CC       ALSO FUNCTION IN THE REGULATION OF MEMBRANE BIOGENESIS.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein. Mitochondrion
CC       (Potential).
CC   -!- SIMILARITY: BELONGS TO THE GPAT / DAPAT FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U61247; AAB04870.1; -.
DR   WormPep; F08F3.5; CE00946.
DR   InterPro; IPR002123; Acyltransferase.
DR   Pfam; PF01553; Acyltransferase; 1.
KW   Phosphoglycerol biosynthesis; Transferase; Acyltransferase;
KW   Transmembrane; Mitochondrion; Transit peptide.
FT   TRANSIT       1      ?       MITOCHONDRION (POTENTIAL).
FT   CHAIN         ?    716       PROBABLE GLYCEROL 3 PHOSPHATE
FT                                ACYLTRANSFERASE.
FT   TRANSMEM    409    425       POTENTIAL.
SQ   SEQUENCE   716 AA;  82071 MW;  E0A36A4A86FC136D CRC64;

  Query Match             100.0%;  Score 45;  DB 1;  Length 716;
  Best Local Similarity   100.0%;  Pred. No. 3e-;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 WXXWHF 6
             ||||||
Db     [** WXXWHF ***

RESULT 15
LYSA_DROME
ID   LYSA_DROME     STANDARD;      PRT;   140 AA.
AC   P41567; P[illegible]4; Q9W[illegible]; Q9W[illegible];
DT   01-OCT-1994 (Rel. 30, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Lysozyme A/C/F precursor (EC 3.2.1.17) (1,4-beta-N-acetylmuramidase
DE   A/C).
GN   LYSA AND (LYSC OR CG9118) AND (LYSD OR CG9119).
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC   Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC   Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Canton-S;
RX   MEDLINE; 94211204; PubMed-8161165;
RA   Daffre S., Kylsten P., Samakovlis C., Hultmark D.;
RT   "The lysozyme locus in Drosophila melanogaster: an expanded gene
RT   family adapted for expression in the digestive tract.";
RL   Mol. Gen. Genet. 242:152-162(1994).

RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Berkeley;
RX   MEDLINE=20530660; PubMed-10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., Mcleod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
CC   -!- FUNCTION: UNLIKELY TO PLAY AN ACTIVE ROLE IN THE HUMORAL IMMUNE
CC       DEFENCE. MAY HAVE A FUNCTION IN THE DIGESTION OF BACTERIA IN THE
CC       FOOD.
CC   -!- CATALYTIC ACTIVITY: Hydrolysis of the 1,4-beta linkages between N-
CC       acetyl-D-glucosamine and N-acetylmuramic acid in peptidoglycan
CC       heteropolymers of the prokaryotes cell walls.
CC   -!- TISSUE SPECIFICITY: FOUND IN THE MIDGUT.
CC   -!- DEVELOPMENTAL STAGE: MAXIMAL EXPRESSION IS FOUND DURING THE THIRD
CC       LARVAL INSTAR, IT DROPS TO BECOME UNDETECTABLE IN THE LATE PUPAL
CC       STAGE. THE EXPRESSION IN ADULTS IS SIMILAR TO THAT OF FIRST AND
CC       SECOND LARVAL INSTARS.
CC   -!- SIMILARITY: BELONGS TO FAMILY [illegible] OF GLYCOSYL HYDROLASES.
CC   -!- CAUTION: Ref.2 sequence differs from that shown due to erroneous
CC       gene model prediction.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z22125; CAA80225.1; -.
DR   EMBL; Z21116; CAA80228.1; -.
DR   EMBL; X58382; CAA41272.1; -.
DR   EMBL; AE003470; AAF47449.1; ALT_SEQ.
DR   PMPD; AP003470; AAF47450.1; -.
DR   PMPD; AP003470; NOT_ANNOTATED_CDS.
DR   PIR; S20914; S20914.
DR   PIR; S32651; S32651.
```

```
DR   PIR; S32726; S32726.
DR   HSSP; P00695; 1LZ5.
DR   FlyBase; FBgn0011201; LysA.
DR   FlyBase; FBgn0004426; LysC.
DR   FlyBase; FBgn0004427; LysD.
DR   InterPro; IPR001916; GH_22.
DR   Pfam; PF00062; lys; 1.
DR   PRINTS; PR00135; LYZLACT.
DR   SMART; SM00263; LYZ1; 1.
DR   PROSITE; PS00128; LACTALBUMIN_LYSOZYME; 1.
KW   Hydrolase; Glycosidase; Bacteriolytic enzyme; Signal;
KW   Multigene family.
FT   SIGNAL        1     18       BY SIMILARITY.
FT   CHAIN        19    140       LYSOZYME A/C/D.
FT   DISULFID     24    139       BY SIMILARITY.
FT   DISULFID     45    129       BY SIMILARITY.
FT   DISULFID     80     96       BY SIMILARITY.
FT   DISULFID     92    110       BY SIMILARITY.
FT   ACT_SITE     50     50       BY SIMILARITY.
FT   ACT_SITE     68     68       BY SIMILARITY.
FT   CONFLICT     41     42       AR -> NK (IN REF. 1).
SQ   SEQUENCE   140 AA;  15635 MW;  75C24CA6F85DF903 CRC64;

  Query Match             92.1%;  Score 35;  DB 1;  Length 140;
  Best Local Similarity   50.0%;  Pred. No. 56;
  Matches    3;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 WXXWHF 6
           |::||:
Db     123 WSTWHY 128


Search completed: February 20, 2003, 10:48:03
Job time : 7.66667 secs
```

GenCore version 5.1.3

OM protein - protein search, using sw model

Run on: February 21, 2003, 10:45:1[illegible] ; Search Time [illegible] Seconds
(without alignments)
52.994 Million cell updates/sec

Title: US-09-912-414-11
Perfect score: 38
Sequence: 1 WXXWHF 6

Scoring table: BLOSUM62X
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_21:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	38	100.0	187	16	Q9CLS4	Q9cls4 pasteurella
2	38	100.0	190	4	Q96N05	Q96n05 homo sapien
3	38	100.0	201	10	Q9C600	Q9c600 arabidopsis
4	38	100.0	202	10	Q9AXY3	Q9axy3 brassica ca
5	38	100.0	213	10	Q9AXY5	Q9axy5 brassica na
6	38	100.0	214	10	Q9AXY2	Q9axy2 brassica ca
7	38	100.0	219	10	Q94K47	Q94k47 arabidopsis
8	38	100.0	222	10	Q9AXY4	Q9axy4 brassica ol
9	38	100.0	230	16	O53604	O53604 mycobacteri
10	38	100.0	241	17	O59123	O59123 pyrococcus
11	38	100.0	256	10	Q9FIN2	Q9fin2 arabidopsis
12	38	100.0	260	16	O83250	O83250 treponema p
13	38	100.0	289	17	Q8TRL3	Q8trl3 methanosarc
14	38	100.0	293	16	Q9I065	Q9i065 pseudomonas
15	38	100.0	310	2	Q9EXF1	Q9exf1 listeria mo
16	38	100.0	314	16	Q9L120	Q9l120 streptomyce
17	38	100.0	381	2	Q93TQ4	Q93tq4 agrobacteri
18	38	100.0	386	16	Q8UH89	Q8uh89 agrobacteri
19	38	100.0	394	10	Q9AWR0	Q9awr0 quillardia
20	38	100.0	403	16	Q8YFS6	Q8yfs6 brucella me
21	38	100.0	415	16	Q8RD64	Q8rd64 thermoanaer
22	38	100.0	421	10	Q9CEU4	Q9ceu4 chlamydomon
23	38	100.0	423	16	Q927H4	Q927h4 listeria in
24	38	100.0	423	16	Q8Y412	Q8y412 listeria mo
25	38	100.0	424	10	Q48663	Q48663 chlamydomon
26	38	100.0	452	16	Q8ZL19	Q8zl19 salmonella
27	38	100.0	452	16	Q8Z2K1	Q8z2k1 salmonella
28	38	100.0	454	10	Q9ZRW0	Q9zrw0 cicer ariet
29	38	100.0	462	16	Q8X5T1	Q8x5t1 escherichia
30	38	100.0	472	16	Q8YMN8	Q8ymn8 anabaena sp
31	38	100.0	473	5	Q9BHU7	Q9bhu7 leishmania
32	38	100.0	473	11	Q8J2S6	Q8j2s6 sus musculu
33	38	100.0	475	17	Q9HSR8	Q9hsr8 halobacteri
34	38	100.0	485	10	Q80401	Q80401 oryza sativ
35	38	100.0	487	5	Q45012	Q45012 caenorhabdi
36	38	100.0	488	16	Q9A5V0	Q9a5v0 caulobacter
37	38	100.0	492	16	Q8YUZ1	Q8yuz1 anabaena sp
38	38	100.0	494	17	Q974C6	Q974c6 sulfolobus
39	38	100.0	497	3	O94514	O94514 schizosacch
40	38	100.0	499	10	Q9LTV0	Q9ltv0 arabidopsis
41	38	100.0	508	3	Q9P7S7	Q9p7s7 schizosacch
42	38	100.0	508	17	Q97WR4	Q97wr4 sulfolobus
43	38	100.0	510	5	Q9U5W4	Q9u5w4 drosophila
44	38	100.0	510	16	Q8XQC5	Q8xqc5 ralstonia s
45	38	100.0	511	5	Q9VM69	Q9vm69 drosophila

ALIGNMENTS

RESULT 1
Q9CLS4
ID Q9CLS4 PRELIMINARY; PRT; 187 AA.
AC Q9CLS4;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01 JUN 2001 (TrEMBLrel. 17, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Hypothetical protein PM1135.
GN PM1135.
OS Pasteurella multocida.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Pasteurella.
OX NCBI_TaxID=747;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PM70;
RX MEDLINE=21145866; PubMed=11248100;
RA May B.J., Zhang Q., Li L.L., Paustian M.L., Whittam T.S., Kapur V.;
RT "Complete genomic sequence of Pasteurella multocida Pm70.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).
CC -!- SIMILARITY: BELONGS TO THE TETR/ACRR FAMILY OF TRANSCRIPTIONAL
CC REGULATORS.
DR EMBL; AE006154; AAK03219.1; -.
DR InterPro; IPR001647; HTH_TetR.
DR Pfam; PF00440; tetR; 1.
DR PRINTS; PR00455; HTHTETR.
KW DNA-binding, Hypothetical protein; Transcription regulation;
KW Complete proteome.
SQ SEQUENCE 187 AA, 22115 MW, [illegible] CRC64;

Query Match 100.0%, Score 38, DB 16, Length 187;
Best Local Similarity 66.7%; Pred. No. 1.9e+02;
Matches 4; Conservative 2, Mismatches 0; Indels 0; Gaps 0;

Qy 1 WXXWHF 6
|::|||
Db 89 WNIWHF 94

```
RESULT 2
Q96NC5
ID   Q96NC5      PRELIMINARY;      PRT;   190 AA.
AC   Q96NC5;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   cDNA FLJ31603 fis, clone NT.BLAS 2[illegible].
OS   Homo sapiens (Human).
OC   Eukaryota, Metazoa, Chordata, Craniata, Vertebrata, Euteleostomi,
OC   Mammalia, Eutheria, Primates, Catarrhini, Hominidae, Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Nishiyama K., Wajatsuma M., Maeda Y., Maeda H., Yokoi T., Kodaira H.,
RA   Furuya T., Takahashi M., Kikkawa F., Omura Y., Abe K., Kamihara K.,
RA   Katsuta N., Sato K., Tanikawa M., Yamazaki M., Sugiyama T., Ito E.,
RA   Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto Y., Isogai Y.,
RA   Kawai Hie Y., Saito K., Nishikawa T., Kimura K., Yamashita H.,
RA   Matsuo K., Nakamura Y., Ohkubo M., Oji Y., Murakawa K.,
RA   Kanehori K., Takahashi Fujii A., Oshima A., Sugiyama A., Kawakami B.,
RA   Suzuki Y., Sugano S., Nujiyori K., Muraki Y., Nagai K., Isogai T.,
RT   "NEDO human cDNA sequencing project.";
RL   Submitted (OCT-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AK056171; BAB71110.1; -.
SQ   SEQUENCE   190 AA;  21018 MW;  B4AF[illegible]4A[illegible]B 75C64;

  Query Match             100.0%;  Score 38;  DB 4;  Length 190;
  Best Local Similarity   66.7%;  Pred. No. 1.6e+02;
  Matches    4;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 WXXWHF 6
           |::|||
Db      82 WGKWHF 87


RESULT 3
Q9C600
ID   Q9C600      PRELIMINARY;      PRT;   201 AA.
AC   Q9C600;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN 2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN 2001 (TrEMBLrel. 17, Last annotation update)
DE   Omega 6 reductase (Fragment).
GN   FAD6.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota, Viridiplantae, Streptophyta, Embryophyta, Tracheophyta,
OC   Spermatophyta; Magnoliophyta, eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID 3702.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. COLUMBIA;
RA   Fourmann M., Froget N., Brunel D.;
RT   "Amplified consensus gene markers: Tools designing for a genetic map
RT   of Arabidopsis known function genes in Brassica.";
RL   Submitted (JAN 2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF229392; AAK00664.1; -.
DR   InterPro; IPR001225; FA_desaturase.
DR   Pfam; PF00487; FA_desaturase; 1.
DR   ProDom; PD001081; FA_desaturase; 1.
FT   NON_TER       1       1
FT   NON_TER     201     201
SQ   SEQUENCE   201 AA;  23084 MW;  3C91FF[illegible]FF[illegible]CF64;

  Query Match             100.0%;  Score 38;  DB 10;  Length 201;
  Best Local Similarity   66.7%;  Pred. No. 2e+02;
  Matches    4;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 WXXWHF 6
           |::|||
Db     156 WVNWHF 161


RESULT 4
Q9AXY3
ID   Q9AXY3      PRELIMINARY;      PRT;   202 AA.
AC   Q9AXY3;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   [illegible] (TrEMBLrel. 17, Last annotation update)
DE   Omega 6 reductase (Fragment).
GN   FAD6 BR-1.
OS   Brassica campestris (Field mustard).
OC   Eukaryota, Viridiplantae, Streptophyta, Embryophyta, Tracheophyta;
OC   Spermatophyta, Magnoliophyta, eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales, Brassicaceae; Brassica.
OX   NCBI_TaxID=3711;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. BP80;
RA   Fourmann M., Froger N., Brunel D.;
RT   "Amplified consensus gene markers: Tools designing for a genetic map
RT   of Arabidopsis known function genes in Brassica.";
RL   Submitted (JAN 2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF229390; AAK00662.1; -.
DR   InterPro; IPR001225; FA_desaturase.
DR   Pfam; PF00487; FA_desaturase; 1.
DR   ProDom; PD001081; FA_desaturase; 1.
FT   NON_TER       1       1
FT   NON_TER     202     202
SQ   SEQUENCE   202 AA;  23161 MW;  A2AB4[illegible]AAB1E[illegible] CRC64;

  Query Match             100.0%;  Score 38;  DB 10;  Length 202;
  Best Local Similarity   66.7%;  Pred. No. 2e+02;
  Matches    4;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 WXXWHF 6
           |::|||
Db     157 WVNWHF 162


RESULT 5
Q9AXY5
ID   Q9AXY5      PRELIMINARY;      PRT;   213 AA.
AC   Q9AXY5;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN 2001 (TrEMBLrel. 17, Last sequence update)
DT   01 JUN 2001 (TrEMBLrel. 17, Last annotation update)
DE   Omega 6 reductase (Fragment).
GN   FAD6-BN-2.
OS   Brassica napus (Rape).
OC   Eukaryota, Viridiplantae, Streptophyta; Embryophyta, Tracheophyta;
OC   Spermatophyta, Magnoliophyta, eudicotyledons, core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Brassica.
OX   NCBI TaxID=3708;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. STELLAR;
RA   Fourmann M., Froger N., Brunel D.;
RT   "Amplified consensus gene markers: Tools designing for a genetic map
RT   of Arabidopsis known function genes in Brassica.";
RL   Submitted (JAN 2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF229388; AAK00660.1; -.
DR   InterPro; IPR001225; FA_desaturase.
DR   Pfam; PF00487; FA_desaturase; 1.
DR   ProDom; PD001081; FA_desaturase; 2.
FT   NON_TER       1       1
FT   NON_TER     213     213
SQ   SEQUENCE   213 AA;  24771 MW;  D40B4B98ADBDC06 CRC64;

  Query Match             100.0%;  Score 38;  DB 10;  Length 213;
  Best Local Similarity   66.7%;  Pred. No. 2.1e+02;
  Matches    4;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 WXXWHF 6
```

```
          |::|||
Db    154 WVNWHF 159


RESULT 6
Q9AXY2
ID   Q9AXY2      PRELIMINARY;      PRT;   216 AA.
AC   Q9AXY2;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Omega 6 reductase (Fragment).
GN   FAD6-BN-1.
OS   Brassica napus (Rape).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Brassica.
OX   NCBI_TaxID=3708;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. STELLAR;
RA   Fourmann M., Froger N., Brunel D.;
RT   "Amplified consensus gene markers. Tools designing for a genetic map
RT   of Arabidopsis-known-function genes in Brassica.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF229391; AAK00663.1; -.
DR   InterPro; IPR001225; FA_desaturase.
DR   Pfam; PF00487; FA_desaturase; 1.
DR   ProDom; PD001081; FA_desaturase; 2.
FT   NON_TER       1      1
FT   NON_TER     216    216
SQ   SEQUENCE   216 AA;  25110 MW;  EFFCEEDBB1B0E3E4 CRC64;

  Query Match             100.0%;  Score 38;  DB 10;  Length 216;
  Best Local Similarity   66.7%;  Pred. No. 2.1e+02;
  Matches    4;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 WXXWHF 6
           |::|||
Db     157 WVNWHF 162


RESULT 7
Q94K47
ID   Q94K47      PRELIMINARY;      PRT;   219 AA.
AC   Q94K47;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Putative 6-phosphogluconolactonase
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Yamada K., Lin S.X., Sakano H., Pham P.K., Park J., Chung M.K.,
RA   Goldsmith A.D., Lee J.M., Quach H.L., Tang C.C., Toriumi M., Yu G.,
RA   Bowser L., Carninci P., Chen H., Cheuk R., Hayashizaki Y., Ishida J.,
RA   Jones T., Kamiya A., Karlin-Neumann G., Kawai J., Kim C., Koesema E.,
RA   Lam B., Lin J., Meyers M.C., Miranda M., Narusaka M., Nguyen M.,
RA   Palm C.J., Sakurai T., Satou M., Seki M., Shinn P., Southwick A.,
RA   Shinozaki K., Davis R.W., Ecker J.R., Theologis A.;
RT   "Full Length cDNA of gene K16H17.11/AT5g24400 (GI:10177901).";
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: CONTAINS 1 WD REPEAT (TRP-ASP DOMAIN).
DR   EMBL; AF370305; AAK44120.1; -.
DR   TIGRFAMs; TIGR01198; pgl; 1.
KW   Repeat; WD repeat.
SQ   SEQUENCE   219 AA;  24542 MW;  1994A723D2434B8A CRC64;

  Query Match             100.0%;  Score 38;  DB 10;  Length 219;
  Best Local Similarity   66.7%;  Pred. No. 2.1e+02;
  Matches    4;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 WXXWHF 6
           |::|||
Db    72 WARWHF 77


RESULT 8
Q9AXY4
ID   Q9AXY4      PRELIMINARY;      PRT;   222 AA.
AC   Q9AXY4;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Omega 6 reductase (Fragment).
GN   FAD6-BO-1.
OS   Brassica oleracea (Cauliflower).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Brassica.
OX   NCBI_TaxID=3712;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. RAPID CYCLING;
RA   Fourmann M., Froger N., Brunel D.;
RT   "Amplified consensus gene markers. Tools designing for a genetic map
RT   of Arabidopsis-known-function genes in Brassica.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF229389; AAK00661.1; -.
DR   InterPro; IPR001225; FA_desaturase.
DR   Pfam; PF00487; FA_desaturase; 1.
DR   ProDom; PD001081; FA_desaturase; 2.
FT   NON_TER       1      1
FT   NON_TER     222    222
SQ   SEQUENCE   222 AA;  25911 MW;  A4B1AFA4B04B7D4B CRC64;

  Query Match             100.0%;  Score 38;  DB 10;  Length 222;
  Best Local Similarity   66.7%;  Pred. No. 2.2e+02;
  Matches    4;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 WXXWHF 6
           |::|||
Db    163 WVNWHF 168


RESULT 9
O53604
ID   O53604      PRELIMINARY;      PRT;   230 AA.
AC   O53604;
DT   01-JUN-1998 (TrEMBLrel. 06, Created)
DT   01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Hypothetical protein Rv0059.
GN   Rv0059 OR MTV030.02 OR MT0065.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Firmicutes; Actinobacteria, Actinobacteridae;
OC   Actinomycetales, Corynebacterineae, Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1773;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37RV;
RX   MEDLINE=98295987; PubMed=9634230;
RA   Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA   Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaia F.,
RA   Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA   Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA   Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA   Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA   Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT   "Deciphering the biology of Mycobacterium tuberculosis from the
RT   complete genome sequence.";
```

```
RL   Nature 393:537-544(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CDC 1551 / [illegible];
RA   Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA   Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA   Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA   Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA   Bishai W.;
RT   "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT   laboratory strains.";
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL021428; CAA16240.1; -.
DR   EMBL; AE006919; AAK41267.1; -.
DR   TIGR; MT0065; -.
DR   TubercuList; Rv0059; -.
KW   Hypothetical protein, Complete proteome.
SQ   SEQUENCE   230 AA;  25664 MW;  1A109D54B1E8E0AE CRC64;

  Query Match           100.0%;  Score 38;  DB 17;  Length 230;
  Best Local Similarity   66.7%;  Pred. No. 2.3e+02;
  Matches    4;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 WXXWHF 6
             |::|||
Db      26 WIVWHF 31


RESULT 10
O59123
ID   O59123      PRELIMINARY;      PRT;   241 AA.
AC   O59123;
DT   01-AUG-1998 (TrEMBLrel. 07, Created)
DT   01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 16, Last annotation update)
DE   Hypothetical protein PH1454.
GN   PH1454.
OS   Pyrococcus horikoshii
OC   Archaea; Euryarchaeota, Thermococci, Thermococcales, Thermococcaceae,
OC   Pyrococcus.
OX   NCBI_TaxID=53953;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=OT3;
RX   MEDLINE 98344137; PubMed=9679194;
RA   Kawarabayasi Y., Sawada M., Horikawa H., Haikawa Y., Hino Y.,
RA   Yamamoto S., Sekine M., Baba S., Kosugi H., Hosoyama A., Nagai Y.,
RA   Sakai M., Ogura K., Otakka R., Nakazawa H., Takamiya M., Ohfuku Y.,
RA   Funahashi T., Tanaka T., Kudoh Y., Yamazaki J., Kushida N., Oguchi A.,
RA   Aoki K., Yoshizawa T., Nakamura Y., Robb F.T., Horikoshi K.,
RA   Masuchi Y., Shizuya H., Kikuchi H.;
RT   "Complete sequence and gene organization of the genome of a hyper-
RT   thermophilic archaebacterium, Pyrococcus horikoshii OT3.";
RL   DNA Res. 5:55-76(1998).
DR   EMBL; AP000006; BAA30041.1; -.
KW   Hypothetical protein, Complete proteome.
SQ   SEQUENCE   241 AA;  25654 MW;  A7977D7FA1F79919 CRC64;

  Query Match           100.0%;  Score 38;  DB 17;  Length 241;
  Best Local Similarity   66.7%;  Pred. No. 2.3e+02;
  Matches    4;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 WXXWHF 6
             |::|||
Db     193 WQTWHF 198


RESULT 11
Q9FIN2
ID   Q9FIN2      PRELIMINARY;      PRT;   256 AA.
AC   Q9FIN2;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   6-phosphogluconolactonase like protein.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota, Viridiplantae, Streptophyta, Embryophyta, Tracheophyta;
OC   Spermatophyta, Magnoliophyta, Mesangiospermae, core eudicots, Rosidae;
OC   eurosids II; Brassicales, Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=COLUMBIA;
RX   MEDLINE 20150233; PubMed=10048488;
RA   Arioli T., Peng L., Betzner A.S., Burn J., Wittke W., Herthum K.,
RA   Tabata S.;
RT   "Structural analysis of Arabidopsis thaliana chromosome 5. VIII.
RT   Sequence features of the regions of 1,081,958 bp covered by seventeen
RT   physically assigned P1 and TAC clones.";
RL   DNA Res. 5:373-391(1998).
DR   EMBL; AB016884; BAB11633.1; -.
DR   InterPro; IPR000457; Glucosamine_iso.
DR   Pfam; PF01182; Glucosamine_iso; 1.
DR   TIGRFAMs; TIGR01198; pgl; 1.
KW   6-PGL; 3D-structure; Complete proteome.
SQ   SEQUENCE   256 AA;  28193 MW;  DB5601245133BE56 CRC64;

  Query Match           100.0%;  Score 38;  DB 16;  Length 256;
  Best Local Similarity   66.7%;  Pred. No. 2.5e+02;
  Matches    4;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 WXXWHF 6
             |::|||
Db      72 WARWHF 77


RESULT 12
O83250
ID   O83250      PRELIMINARY;      PRT;   260 AA.
AC   O83250;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   D,D carboxypeptidase, putative.
GN   TP0221.
OS   Treponema pallidum.
OC   Bacteria; Spirochaetales; Spirochaetaceae; Treponema.
OX   NCBI_TaxID=160;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NICHOLS;
RX   MEDLINE 98317770; PubMed=9688576;
RA   Fraser C.M., Norris S.J., Weinstock G.M., White O., Sutton G.G.,
RA   Dodson R., Gwinn M., Hickey E.K., Clayton R., Ketchum K.A.,
RA   Sodergren E., Hardham J.M., McLeod M.P., Salzberg S., Peterson J.,
RA   Khalak H., Richardson D., Howell J.K., Chidambaram M., Utterback T.,
RA   McDonald L., Artiach P., Bowman C., Cotton M.D., Fujii C., Garland S.,
RA   Hatch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O.,
RA   Venter J.C.;
RT   "Complete genome sequence of Treponema pallidum, the syphilis
RT   spirochete.";
RL   Science 281:375-388(1998).
DR   EMBL; AE001203; AAC65208.1; -.
DR   TIGR; TP0221; -.
DR   InterPro; IPR000561; EGF-like.
DR   InterPro; IPR003709; VanY.
DR   Pfam; PF02557; VanY; 1.
DR   PROSITE; PS00022; EGF_1; UNKNOWN_1.
KW   CarboxypeptidaseD; Complete proteome.
SQ   SEQUENCE   260 AA;  29193 MW;  DB5601245133BE56 CRC64;

  Query Match           100.0%;  Score 38;  DB 16;  Length 260;
  Best Local Similarity   66.7%;  Pred. No. 2.5e+02;
  Matches    4;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 WXXWHF 6
             |::|||
```

```
Db    220 WEPWHF 225


RESULT 13
Q8TRL3
ID   Q8TRL3      PRELIMINARY;      PRT;   293 AA.
AC   Q8TRL3;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Hypothetical protein MA1162.
GN   MA1162.
OS   Methanosarcina acetivorans.
OC   Archaea; Euryarchaeota; Methanomicrobi; Methanosarcinales;
OC   Methanosarcinaceae; Methanosarcina.
OX   NCBI_TaxID=2214;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C2A / ATCC 35395 / DSM 2834;
RX   MEDLINE=21929760; PubMed=11932238;
RA   Galagan J.E., Nusbaum C., Roy A., Endrizzi M.G., Macdonald P.,
RA   FitzHugh W., Calvo S., Engels R., Smirnov S., Atnoor D., Brown A.,
RA   Allen N., Naylor J., Stange-Thomann N., DeArellano K., Johnson R.,
RA   Linton L., McEwan P., McKernan K., Talamas J., Tirrell A., Ye W.,
RA   Zimmer A., Barber R.D., Cann I., Graham D.E., Grahame D.A., Guss A.M.,
RA   Hedderich R., Ingram-Smith C., Kuettner H.C., Krzycki J.A.,
RA   Leigh J.A., Li W., Liu J., Mukhopadhyay B., Reeve J.N., Smith K.,
RA   Springer T.A., Umayam L.A., White O., White R.H., de Macario E.C.,
RA   Ferry J.G., Jarrell K.F., Jing H., Macario A.J.L., Paulsen I.,
RA   Pritchett M., Sowers K.R., Swanson R.V., Zinder S.H., Lander E.,
RA   Metcalf W.W., Birren B.;
RT   "The genome of Methanosarcina acetivorans reveals extensive metabolic
RT   and physiological diversity.";
RL   Genome Res. 12:532-542(2002).
DR   EMBL; AE010783; AAM04583.1; -.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   293 AA;  33289 MW;  B4221A0BEA52BBA6 CRC64;

  Query Match             100.0%;  Score 38;  DB 17;  Length 293;
  Best Local Similarity   66.7%;  Pred. No. 2.9e+02;
  Matches    4;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 WXXWHF 6
           |::|||
Db     168 WSLWHF 173


RESULT 14
Q9I065
ID   Q9I065      PRELIMINARY;      PRT;   293 AA.
AC   Q9I065;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Hypothetical protein PA2778.
GN   PA2778
OS   Pseudomonas aeruginosa
OC   Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC   Pseudomonas.
OX   NCBI_TaxID=287;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 15692 / PAO1;
RX   MEDLINE=20437337; PubMed=10984043;
RA   Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warrener P.,
RA   Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA   Garber R.L., Goltry L., Tolentino E., Westbrock-Wadman S., Yuan Y.,
RA   Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA   Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA   Reizer J., Saier M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT   "Complete genome sequence of Pseudomonas aeruginosa PAO1, an
RT   opportunistic pathogen.";
RL   Nature 406:959-964(2000).
DR   EMBL; AE004795; AAG06166.1; -.
DR   InterPro; IPR005074; Peptidase_C39.
DR   InterPro; IPR001440; TPR.
DR   Pfam; PF03412; Peptidase_C39; 1.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   293 AA;  33116 MW;  [illegible] CRC64;

  Query Match             100.0%;  Score 38;  DB 16;  Length 293;
  Best Local Similarity   66.7%;  Pred. No. 2.9e+02;
  Matches    4;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 WXXWHF 6
           |::|||
Db     101 WPRWHF 106


RESULT 15
Q9EXF1
ID   Q9EXF1      PRELIMINARY;      PRT;   310 AA.
AC   Q9EXF1;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Putative phosphotransferase D-arabitol-specific component IIC
DE   (Fragment).
GN   ATLC.
OS   Listeria monocytogenes.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC   Listeriaceae; Listeria.
OX   NCBI_TaxID=1639.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=LO28;
RX   MEDLINE=21212979; PubMed=11316371;
RA   Saklani-Jusforgues H., Fontan E., Goossens P.L.;
RT   "Characterisation of a Listeria monocytogenes mutant deficient in D-
RT   arabitol fermentation.";
RL   Res. Microbiol. 152:175-177(2001).
DR   EMBL; AJ292552; CAC20641.1; -.
DR   InterPro; IPR004703, Gal_spec_IIC.
DR   Pfam; PF03611, EIIC-GAT; 1.
KW   Transferase.
FT   NON_TER       1      1
FT   NON_TER     310    310
SQ   SEQUENCE   310 AA;  33711 MW;  A2F9FE9CBC88DACF CRC64;

  Query Match             100.0%;  Score 38;  DB 2;  Length 310;
  Best Local Similarity   66.7%;  Pred. No. 3e+02;
  Matches    4;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 WXXWHF 6
           |::|||
Db      18 WNYWHF 23


Search completed. February 20, 2003, 10.43.22
Job time : 25.3333 secs
```

GenCore version 5.1.3

OM protein - protein search, using sw model

Run on: February 20, 2003, 10:35:40 ; Search Time 29.3333 Seconds
(without alignments)
27.256 Million cell updates/sec

Title: US-09-912-414-11
Perfect score: 38
Sequence: 1 WXXWHF 6

Scoring table: BLOSUM62DX
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	38	100.0	6	21	AAB01492	Peptide which bind
2	38	100.0	6	21	AAB01499	Peptide which bind
3	38	100.0	6	21	AAB01505	Peptide which bind
4	38	100.0	6	21	AAB01506	Peptide which bind
5	38	100.0	36	22	ABB40893	Peptide #8399 enco
6	38	100.0	36	22	AAM61753	Human brain expres
7	38	100.0	36	22	AAM74549	Human bone marrow
8	38	100.0	36	22	AAM84663	Peptide #8700 enco
9	38	100.0	43	22	AAU14683	Novel bone marrow
10	38	100.0	54	23	ABG64892	Human albumin fusi
11	38	100.0	53	23	AAE21233	Human gene 18 enco
12	38	100.0	57	22	AAU66178	Propionibacterium
13	38	100.0	85	23	ABF07322	Human ORFX protein
14	38	100.0	87	22	AAU56391	Propionibacterium
15	38	100.0	96	22	AAU14777	Novel bone marrow
16	38	100.0	110	23	ABF06130	Human ORFX protein
17	38	100.0	173	22	AAG80884	Lipid modification
18	38	100.0	173	22	AAG83336	Fatty lipid met
19	38	100.0	175	22	ABB03211	Human musculoskele
20	38	100.0	175	22	AAM42372	Human polypeptide
21	38	100.0	256	21	AAG09463	Arabidopsis thalia
22	38	100.0	256	21	AAG44075	Arabidopsis thalia
23	38	100.0	256	21	AAG48293	Arabidopsis thalia
24	38	100.0	258	22	AAG93113	C glutamicum prote
25	38	100.0	306	21	AAG05810	Arabidopsis thalia
26	38	100.0	306	21	AAG47194	Arabidopsis thalia
27	38	100.0	324	21	AAG50316	Arabidopsis thalia
28	38	100.0	325	21	AAG09462	Arabidopsis thalia
29	38	100.0	325	21	AAG44074	Arabidopsis thalia
30	38	100.0	325	21	AAG48292	Arabidopsis thalia
31	38	100.0	338	21	AAG05809	Arabidopsis thalia
32	38	100.0	338	21	AAG47193	Arabidopsis thalia
33	38	100.0	376	19	AAW46297	Rhizobium meliloti
34	38	100.0	377	19	AAW46298	Brucella abortus D
35	38	100.0	418	15	AAB60501	Linoleic-acid-desa
36	38	100.0	423	21	AAG50315	Arabidopsis thalia
37	38	100.0	423	23	ABP43693	Listeria monocyto
38	38	100.0	424	22	AAE13431	Soybean microsomal
39	38	100.0	427	22	AAU20491	Human secreted pro
40	38	100.0	427	22	AAU21839	Novel human neopla
41	38	100.0	448	21	AAG50308	Arabidopsis thalia
42	38	100.0	448	21	AAG47192	Arabidopsis thalia
43	38	100.0	448	24	ABB03179	Herbicidally activ
44	38	100.0	452	21	AAG44073	Arabidopsis thalia
45	38	100.0	466	22	AAB05611	Human protein sequ

ALIGNMENTS

RESULT 1
AAB01492
ID AAB01492 standard, peptide, 6 AA.
XX
AC AAB01492;
XX
DT 08-NOV-2000 (first entry)
XX
DE Peptide which binds to transcription factor E2F 1 DNA binding domain.
XX
KW DNA binding, transcription factor, E2F, E2F 1, cell cycle, DP-1;
KW activation, transcription, apoptosis, proliferative disorder;
KW psoriasis, restenosis.
XX
OS Synthetic.
XX
PN WO200044771-A1.
XX
PD 03-AUG-2000.
XX
PF 26-JAN-2000; 2000WO-GB00227.
XX
PR 26-JAN-1999; 99GB-0001710.
XX
PA (PROL-) PROLIFIX LTD.
XX
PI Mueller P, Kontermann PE, Montigiani S;
XX
DR WPI; 2000-532806/48.
XX
PT Peptides binding to the DNA binding domain of transcription factor E2F
PT and inhibiting cell cycle progression, useful for the treatment of
PT cancer

```
XX
PS   Claim 6; Page 2; 42pp; English.
XX
CC   Peptides which bind to the DNA binding domain of transcription
CC   factor E2F and inhibit cell cycle progression may be useful as
CC   research agents to investigate the interaction between E2F and DP-1,
CC   or the activation of transcription by E2F-1/DP-1 heterodimers. They
CC   may also be used for inducing apoptosis and/or cell cycle arrest in
CC   a cell, particularly for treatment of cancer or other proliferative
CC   disorders such as psoriasis and restenosis.
XX
SQ   Sequence   6 AA;

  Query Match             100.0%;  Score 38;  DB 21;  Length 6;
  Best Local Similarity   66.7%;  Pred. No. 7.8e+05;
  Matches    4;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 WXXWHF 6
           |::|||
Db       1 WVRWHF 6


RESULT 2
AAB01499
ID   AAB01499 standard; peptide; 6 AA
XX
AC   AAB01499;
XX
DT   08-NOV-2000  (first entry)
XX
DE   Peptide which binds to transcription factor E2F 1 DNA binding domain
XX
KW   DNA binding, transcription factor, E2F, E2F 1, cell cycle, DP 1,
KW   activation; transcription; apoptosis, proliferative disorder,
KW   psoriasis, restenosis.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 2
FT                   /note= "Any amino acid"
FT   Misc difference 3
FT                   /note= "Any amino acid"
XX
PN   WO200044771 A1.
XX
PD   03-AUG-2000.
XX
PF   26 JAN 2000;  2000WO-EB00777
XX
PR   26-JAN 1999;  99GB-0001710
XX
PA   (PROL-) PROLIFIX LTD.
XX
PI   Mueller R;  Kontermann RE;  Montigiani S;
XX
DR   WPI, 2000 532806/48
XX
PT   Peptides binding to the DNA binding domain of transcription factor E2F
PT   and inhibiting cell cycle progression, useful for the treatment of
PT   cancer
XX
PS   Claim 4; Page 9; 42pp; English.
XX
CC   Peptides which bind to the DNA binding domain of transcription
CC   factor E2F and inhibit cell cycle progression may be useful as
CC   research agents to investigate the interaction between E2F and DP-1,
CC   or the activation of transcription by E2F-1/DP-1 heterodimers. They
CC   may also be used for inducing apoptosis and/or cell cycle arrest in
CC   a cell, particularly for treatment of cancer or other proliferative
CC   disorders such as psoriasis and restenosis.
XX
SQ   Sequence   6 AA;

  Query Match             100.0%;  Score 38;  DB 21;  Length 6;
  Best Local Similarity   100.0%;  Pred. No. 7.8e+05;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 WXXWHF 6
           ||||||
Db       1 WXXWHF 6


RESULT 3
AAB01505
ID   AAB01505 standard; peptide; 6 AA.
XX
AC   AAB01505;
XX
DT   08-NOV-2000  (first entry)
XX
DE   Peptide which binds to transcription factor E2F 1 DNA binding domain.
XX
KW   DNA binding, transcription factor, E2F, E2F 1, cell cycle; DP-1;
KW   activation, transcription, apoptosis, proliferative disorder;
KW   psoriasis; restenosis.
XX
OS   Synthetic
XX
PN   WO200044771-A1.
XX
PD   03-AUG-2000.
XX
PF   26-JAN-2000; 2000WO-EB00777.
XX
PR   26-JAN-1999;  99GB-0001710.
XX
PA   (PROL-) PROLIFIX LTD.
XX
PI   Mueller R;  Kontermann RE;  Montigiani S;
XX
DR   WPI, 2000-532806/48.
XX
PT   Peptides binding to the DNA binding domain of transcription factor E2F
PT   and inhibiting cell cycle progression, useful for the treatment of
PT   cancer
XX
PS   Example; Page 26; 42pp; English.
XX
CC   Peptides which bind to the DNA binding domain of transcription
CC   factor E2F and inhibit cell cycle progression may be useful as
CC   research agents to investigate the interaction between E2F and DP-1,
CC   or the activation of transcription by E2F-1/DP-1 heterodimers. They
CC   may also be used for inducing apoptosis and/or cell cycle arrest in
CC   a cell, particularly for treatment of cancer or other proliferative
CC   disorders such as psoriasis and restenosis.
XX
SQ   Sequence   6 AA;

  Query Match             100.0%;  Score 38;  DB 21;  Length 6;
  Best Local Similarity   66.7%;  Pred. No. 7.8e+05;
  Matches    4;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 WXXWHF 6
           |::|||
Db       1 WARWHF 6


RESULT 4
AAB01506
ID   AAB01506 standard; peptide; 6 AA.
XX
AC   AAB01506;
XX
DT   08-NOV-2000  (first entry)
XX
```

DE Peptide which binds to transcription factor E2F-1 DNA binding domain
XX
KW DNA binding, transcription factor, E2F, E2F-1, cell cycle, DP-1,
KW activation; transcription; apoptosis, proliferative disorder,
KW psoriasis; restenosis.
XX
OS Synthetic.
XX
PN WO200044771-A1.
XX
PD 03-AUG-2000.
XX
PF 26-JAN-2000; 2000WO-GB00227.
XX
PR 26-JAN-1999; 99GB-0001710.
XX
PA (PROL-) PROLIFIX LTD.
XX
PI Mueller R, Kontermann PE, Montigiani S;
XX
DR WPI; 2000-532806/48
XX
PT Peptides binding to the DNA binding domain of transcription factor E2F
PT and inhibiting cell cycle progression, useful for the treatment of
PT cancer
XX
PS Example; Page 26; 42pp, English.
XX
CC Peptides which bind to the DNA binding domain of transcription
CC factor E2F and inhibit cell cycle progression may be useful as
CC research agents to investigate the interaction between E2F and DP-1,
CC or the activation of transcription by E2F-1/DP-1 heterodimers. They
CC may also be used for inducing apoptosis and/or cell cycle arrest in
CC a cell, particularly for treatment of cancer or other proliferative
CC disorders such as psoriasis and restenosis
XX
SQ Sequence 6 AA;

Query Match 100.0%; Score 38; DB 21; Length 6;
Best Local Similarity 66.7%; Pred. No. 7.8e+05;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 WXXWHF 6
|::|||
Db 1 WVAWHF 6

RESULT 5
ABB40893
ID ABB40893 standard; Peptide; 36 AA.
XX
AC ABB40893;
XX
DT 04-FEB-2002 (first entry)
XX
DE Peptide #8399 encoded by human foetal liver single exon probe
XX
KW Human; foetal liver; gene expression; single exon nucleic acid probe
XX
OS Homo sapiens.
XX
PN WO200157277-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00669.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-483447/52.
XX
PT Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human fetal liver -
XX
PS Claim 27; SEQ ID NO 33528; 639pp + sequence listing, English.
XX
CC The invention relates to a single exon nucleic acid probe for
CC measuring human gene expression in a sample derived from human foetal
CC liver. The single exon nucleic acid probes may be used for predicting,
CC measuring and displaying gene expression in samples derived from human
CC fetal liver. The present sequence is a peptide encoded by a single exon
CC nucleic acid probe of the invention.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp wipo int/pub/published_pct_sequences
XX
SQ Sequence 36 AA;

Query Match 100.0%; Score 38; DB 22; Length 36;
Best Local Similarity 66.7%; Pred. No. 56;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 WXXWHF 6
|::|||
Db 19 WRAWHF 24

RESULT 6
AAM61753
ID AAM61753 standard; Protein; 36 AA.
XX
AC AAM61753;
XX
DT 01-NOV-2001 (first entry)
XX
DE Human brain expressed single exon probe encoded protein SEQ ID NO: 33858.
XX
KW Human; brain expressed exon; gene expression analysis; probe;
KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
KW epilepsy; cancer.
XX
OS Homo sapiens.
XX
PN WO200157275-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00667.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-483446/52.
XX
PT Single exon nucleic acid probes for analyzing gene expression in human
PT brains -
XX

```
PS   Example 4; SEQ ID NO: 33858; 650pp + Sequence Listing; English.
XX
CC   The present invention provides a number of single exon nucleic acid
CC   probes which are derived from genomic sequences expressed in the human
CC   brain. They can be used to measure gene expression in brain cell samples,
CC   which may enable the diagnosis and improved treatment of nervous system
CC   diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
CC   epilepsy and cancers. The present sequence is a protein encoded by one of
CC   the probes of the invention
XX
SQ   Sequence   36 AA;

  Query Match             100.0%;  Score 38;  DB 22;  Length 36;
  Best Local Similarity   66.7%;  Pred. No. 56;
  Matches    4;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 WXXWHF 6
           |::|||
Db      19 WRAWHF 24


RESULT 7
AAM74549
ID   AAM74549 standard; Protein; 36 AA.
XX
AC   AAM74549;
XX
DT   06-NOV-2001  (first entry)
XX
DE   Human bone marrow expressed protein encoded by SEQ ID NO 33858.
XX
KW   Human; bone marrow expressed exon; gene expression analysis; probe;
KW   microarray; cancer; leukaemia; lymphoma; myeloma.
XX
OS   Homo sapiens.
XX
PN   WO200157276-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US00668.
XX
PR   04-FEB-2000; 2000US-0180311.
PR   26-MAY-2000; 2000US-0207456.
PR   30-JUN-2000; 2000US-0608409.
PR   03-AUG-2000; 2000US-0632366.
PR   21-SEP-2000; 2000US-0234697.
PR   27-SEP-2000; 2000US-0236359.
PR   04-OCT-2000; 2000GB-0024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR   WPI; 2001-488900/53.
XX
PT   Human genome-derived single exon nucleic acid probes useful for
PT   analysing gene expression in human bone marrow
XX
PS   Example 4; SEQ ID NO: 33858; 650pp + Sequence Listing; English.
XX
CC   The present invention provides a number of single exon nucleic acid
CC   probes which are derived from genomic sequences expressed in the human
CC   bone marrow. They can be used to measure gene expression in bone marrow
CC   samples, which may enable the improved diagnosis and treatment of cancers
CC   such as lymphoma, leukaemia and myeloma. The present sequence is a
CC   protein encoded by one of the probes of the invention.
XX
SQ   Sequence   36 AA;

  Query Match             100.0%;  Score 38;  DB 22;  Length 36;
  Best Local Similarity   66.7%;  Pred. No. 56;
  Matches    4;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 WXXWHF 6
           |::|||
Db      19 WRAWHF 24


RESULT 8
AAM34663
ID   AAM34663 standard; Protein; 36 AA.
XX
AC   AAM34663;
XX
DT   17-OCT-2001  (first entry)
XX
DE   Peptide #8700 encoded by probe for measuring placental gene expression.
XX
KW   Probe; microarray; human; placenta; antenatal diagnosis;
KW   genetic disorder.
XX
OS   Homo sapiens.
XX
PN   WO200157272-A2.
XX
PD   09-AUG-2001.
XX
PF   29-JAN-2001; 2001WO-US00663.
XX
PR   04-FEB-2000; 2000US-0180312.
PR   26-MAY-2000; 2000US-0207456.
PR   30-JUN-2000; 2000US-0608408.
PR   03-AUG-2000; 2000US-0632366.
PR   21-SEP-2000; 2000US-0234697.
PR   27-SEP-2000; 2000US-0236359.
PR   04-OCT-2000; 2000GB-0024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR   WPI; 2001-488897/53.
XX
PT   Human genome-derived single exon nucleic acid probes useful for
PT   analysing gene expression in human placenta
XX
PS   Claim 27; SEQ ID NO 34912; 654pp; English.
XX
CC   The present invention relates to single exon nucleic acid probes (CENP:
CC   see AAI31315-AAI57646). The present sequence is a peptide encoded by one
CC   such probe. The probes are useful for producing a microarray for
CC   predicting, measuring and displaying gene expression in samples derived
CC   from human placenta. The probes are useful for antenatal diagnosis of
CC   human genetic disorders.
XX
SQ   Sequence   36 AA;

  Query Match             100.0%;  Score 38;  DB 22;  Length 36;
  Best Local Similarity   66.7%;  Pred. No. 56;
  Matches    4;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 WXXWHF 6
           |::|||
Db      19 WRAWHF 24


RESULT 9
AAU14683
ID   AAU14683 standard; Protein; 10 AA.
XX
AC   AAU14683;
XX
DT   24-OCT-2001  (first entry)
XX
DE   Novel bone marrow polypeptide #80.
```

```
XX
KW   Bone marrow; diagnostic; therapeutic; gene therapy; antigenic;
KW   haematopoiesis; myeloid; lymph cell disorder; tissue regeneration;
KW   wound healing; nutritional supplement; immune disorder;
KW   severe combined immunodeficiency; SCID.
XX
OS   Homo sapiens.
XX
PN   WO200157187-A2.
XX
PD   09-AUG-2001.
XX
PF   05-FEB-2001; 2001WO-US03782.
XX
PR   03-FEB-2000; 2000US-0496914.
PR   20-JUN-2000; 2000US-0598075.
PR   19-JUL-2000; 2000US-0620325.
PR   30-NOV-2000; 2000US-0250683.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Ford JE,  Boyle BT,  Tang YT,  Lin Z,  Asundi V,  Chen P,  Xue AJ;
PI   Ren F,  Drmanac RT;
XX
DR   WPI; 2001-488876/53
DR   N-PSDB; AAQ22988
XX
PT   Nucleic acids encoding bone marrow polypeptides, useful in diagnostic
PT   and gene therapy -
XX
PS   Claim 10; Page 248; 392pp; English.
XX
CC   AAU14602-AAU14794 represent novel bone marrow polypeptides of the
CC   invention. The proteins and corresponding coding sequences may be used
CC   in the prevention, diagnosis and treatment of diseases associated with
CC   inappropriate bone marrow polypeptide expression. For example, to treat
CC   disorders associated with decreased expression by rectifying mutations
CC   or deletions in a patient's genome that affect the activity of the
CC   polypeptides by expressing inactive proteins or to supplement the
CC   patient's own production of the polypeptide. Additionally, the nucleic
CC   acids may be used to produce the polypeptides, by inserting the nucleic
CC   acids into a host cell and culturing the cell to express the protein.
CC   The nucleic acid and its complementary sequences may also be used as DNA
CC   probes in diagnostic assays to detect and quantitate the presence of
CC   similar nucleic acid sequences in samples, and therefore which patients
CC   may be in need of restorative therapy. The proteins may also be used as
CC   antigens in the production of antibodies against bone marrow proteins
CC   and in assays to identify modulators of their expression and activity.
CC   The anti-bone marrow protein antibodies and antagonists may also be used
CC   to down regulate expression and activity. The antibodies may also be used
CC   as diagnostic agents for detecting the presence of the protein in samples
CC   (e.g. by enzyme linked immunosorbant assay (ELISA)). The proteins
CC   may be used to regulate haematopoiesis activity, and consequently in the
CC   treatment of myeloid or lymph cell disorders; in tissue regeneration,
CC   such as wound healing; as a nutritional supplement; and in treatment of
CC   immune disorders such as severe combined immunodeficiency (SCID).
XX
SQ   Sequence   43 AA;

  Query Match             100.0%;  Score 38;  DB 22;  Length 43;
  Best Local Similarity   66.7%;  Pred. No. 67;
  Matches    4;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 WXXWHF 6
           |::|||
Db      14 WLIWHF 19


RESULT 10
ABG64892
ID   ABG64892 standard; Protein; 53 AA.
XX
AC   ABG64892;
XX
DT   27-AUG-2002  (first entry)
XX
DE   Human albumin fusion protein #1567.
XX
KW   Albumin fusion protein; therapeutic protein X; human albumin; HA;
KW   human serum albumin; HSA; cancer; reproductive disorder;
KW   digestive disorder; immune disorder; endocrine disorder;
KW   haematopoietic disorder; neural disorder; connective disorder;
KW   cytostatic; antiinfertility; antiinflammatory; antiulcer;
KW   immunomodulator; anti HIV; antidiabetic; haemostatic; nootropic;
KW   neuroprotective; antiparkinsonian; antimicrobial; neuroleptic;
KW   osteopathic; antiarthritic.
XX
OS   Homo sapiens.
OS   Synthetic.
XX
PN   WO200177137-A1.
XX
PD   18-OCT-2001.
XX
PF   12-APR-2001; 2001WO-US11588.
XX
PR   12-APR-2000; 2000US-229358P.
PR   25-APR-2000; 2000US-199384P.
PR   21-JUL-2000; 2000US-2660?1P
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Haseltine WA;
XX
DR   WPI; 2002-010886/01.
XX
PT   New fusion protein for treating disease e.g. diabetes comprises an
PT   albumin fused to a therapeutic protein -
XX
PS   Claim 1; Page 1582; 2102pp; English.
XX
CC   The present invention relates to albumin fusion proteins comprising a
CC   therapeutic protein X and human albumin (HA, also known as human serum
CC   albumin, HSA). The proteins are useful for treating a disease or
CC   disorder that may be modulated by therapeutic protein X. The albumin
CC   extends the shelf-life of protein X, and may increase its biological
CC   in vitro/in vivo activity. The protein is useful for treating and
CC   diagnosing disorders such as cancer, reproductive disorders, digestive
CC   disorders (e.g. Crohn's disease, ulcerative colitis), immune disorders
CC   (e.g. acquired immunodeficiency syndrome, AIDS), endocrine disorders
CC   (e.g. diabetes), haematopoietic disorders, neural disorders
CC   (e.g. Alzheimer's, Parkinson's, Creutzfeldt-Jacob disease,
CC   encephalomyelitis, meningitis, schizophrenia), and connective disorders
CC   (e.g. osteoporosis, arthritis). ABG63326-ABG65518 represent albumin
CC   fusion proteins of the invention.
XX
SQ   Sequence   53 AA;

  Query Match             100.0%;  Score 38;  DB 23;  Length 53;
  Best Local Similarity   66.7%;  Pred. No. 81;
  Matches    4;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 WXXWHF 6
           |::|||
Db      21 WASWHF 26


RESULT 11
AAE21233
ID   AAE21233 standard; Protein; 53 AA.
XX
AC   AAE21233;
XX
DT   01-JUL-2002  (first entry)
XX
DE   Human gene 18 encoded secreted protein HNNBM45, SEQ ID NO:98.
```

XX
KW Human; secreted protein; immune disorder; antiallergic; antirheumatic;
KW rheumatoid arthritis; breast neoplasia; breast cancer; antiarthritic;
KW neurological disease, Alzheimer's disease, Parkinson's disease, trauma,
KW ischemic epilepsy, encephalitis, cytostatic, hemostasis, anaemia, sepsis,
KW anti-inflammatory; ophthalmological, dermatological, immunostimulatory,
KW immunomodulatory; immunosuppressive; antibacterial; antipsoriatic;
KW gene therapy, autoimmune disease, Huntington's disease, meningitis,
KW demyelinating disease; peripheral neuropathy; congenital malformation;
KW spinal cord injury, peripheral neuropathy, ischaemia, perception,
KW multiple sclerosis, infarction, haemorrhage, schizophrenia, dementia,
KW depression, panic disorder, learning disability, ALS, feeding disorder;
KW hyperproliferative disorder, sleep pattern, cardiovascular disorder,
KW reproductive disorder, digestive system disorder, behavioural disorder.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Peptide 1..31
FT /label= Signal_peptide
FT Protein 32..53
FT /label= Mature_secreted_protein
FT Misc_difference 50
FT /label= Unknown
FT /note= "Encoded by NCA"
XX
PN WO200216390-A1.
XX
PD 28-FEB 2002.
XX
PF 17-JAN-2001; 2001WO-US01435.
XX
PR 18-AUG-2000; 2000US-226282P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Ruben SM, Baker KP, Barash JE, Dillon PJ, Li Y,
PI Moore PA, Wei P, Ebner R, Duan RD, Shi Y, Choi GH, Fikes JD,
PI Ni J;
XX
DR WPI; 2002-304113/34.
DR N-PSDB; AAD33734.
XX
PT An isolated nucleic acid molecule (I) comprising a polynucleotide which
PT encodes a polypeptide useful in the diagnosis and treatment of
PT disorders e.g. immune disorders -
XX
PS Claim 11; Page 504; 534pp; English.
XX
CC AAD33690-AAD33736 represent cDNAs corresponding to 21 human secreted
CC protein genes, and AAE21191-AAE21238 represent the proteins they encode.
CC AAE21236-AAE21237 represent human secreted protein fragments. The genes
CC and their corresponding secreted proteins are useful for preventing,
CC treating or ameliorating medical conditions, e.g., by protein or gene
CC therapy. Pathological conditions can be diagnosed, by detecting the
CC amount of the new protein in a sample or by determining the presence of
CC mutations in the new genes. Specific uses are described for each of the
CC 21 genes, based on the tissues in which they are most highly expressed,
CC and include developing products for the diagnosis or treatment of
CC immune or autoimmune disorders e.g. AIDS (acquired immune deficiency
CC syndrome), asthma, anaemia and rheumatoid arthritis, breast neoplasia
CC and breast cancer, neurological diseases e.g. Alzheimer's disease,
CC Parkinson's disease, Huntington's disease, Tourette syndrome,
CC meningitis, demyelinating disease, peripheral neuropathies, neoplasia,
CC trauma, congenital malformations, spinal cord injuries, toxic
CC neuropathies induced by neurotoxins, peripheral neuropathies, multiple
CC sclerosis, ischaemia and infarction, haemorrhage, schizophrenia, mania,
CC dementia, depression, panic disorder, learning disabilities, ALS,
CC altered behaviour e.g. disorders in feeding, sleep patterns, balance
CC and perception, encephalitis, disorders in cardiovascular, neural/
CC sensory, reproductive and digestive systems, behavioural disorders and
CC hyperproliferative disorder. The present sequence represents a human
CC secreted protein of the invention.

XX
SQ Sequence 53 AA;

Query Match 100.0%; Score 46; DB 23; Length 53;
Best Local Similarity 77.8%; Pred. No. 81;
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 WXXWHF 6
|::|||
Db 21 WACWHF 26

RESULT 12
AAU66178
ID AAU66178 standard; Protein; 57 AA.
XX
AC AAU66178;
XX
DT 27-FEB 2002 (first entry)
XX
DE Propionibacterium acnes immunogenic protein #27074.
XX
KW SAPHO syndrome, synovitis, acne, pustulosis, hyperostosis, osteomyelitis;
KW uveitis; endophthalmitis; bone, joint; central nervous system; ELISA;
KW inflammatory lesion, acne vulgaris, enzyme linked immunosorbent assay;
KW dermatological; osteopathic; neuroprotectant
XX
OS Propionibacterium acnes.
XX
PN WO200181581-A2.
XX
PD 01-NOV-2001.
XX
PF 20-APR-2001; 2001WO-US12865.
XX
PR 21-APR-2000; 2000US-199047P.
PR 06-JUN-2000; 2000US-208841P.
PR 07-JUL-2000; 2000US-216717P.
XX
PA (CORI-) CORIXA CORP.
XX
PI Skeiky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;
PI L'maisonneuve J, Zhang Y, Jen S, Carter D;
XX
DR WPI; 2001-616774/71.
DR N-PSDB; AAS59718.
XX
PT Propionibacterium acnes polypeptides and nucleic acids useful for
PT vaccinating against and diagnosing infections, especially useful for
PT treating acne vulgaris -
XX
PS Example 1; SEQ ID NO 27073; 1060pp; English.
XX
CC Sequences AAU66107-AAU66217 represent Propionibacterium acnes immunogenic
CC polypeptides. The proteins and their associated DNA sequences are used in
CC the treatment, prevention and diagnosis of medical conditions caused by
CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,
CC pustulosis, hyperostosis and osteomyelitis), uveitis and endophthalmitis.
CC P. acnes is also involved in infections of bone, joints and the central
CC nervous system, however it is particularly involved in the inflammatory
CC lesions associated with acne vulgaris. A method for detecting the
CC presence or absence of P. acnes in a patient comprises contacting a
CC sample with a binding agent that binds to the proteins of the invention
CC and determining the amount of bound protein in the sample. The
CC polypeptides may be used as antigens in the production of antibodies
CC specific for P. acnes proteins. These antibodies can be used to
CC immunize the experimental animal and antibody of P. acnes polypeptides and
CC therefore treat P. acnes infections. The antibodies may also be used as
CC diagnostic agents for determining P. acnes presence, for example, by
CC enzyme linked immunosorbent assay (ELISA).
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

```
XX
SQ   Sequence   57 AA;

  Query Match             100.0%;  Score 38;  DB 22;  Length 57;
  Best Local Similarity   66.7%;  Pred. No. 87;
  Matches    4;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 WXXWHF 6
           |::|||
Db       4 WPPWHF 9


RESULT 13
ABP07322
ID   ABP07322 standard; Protein; 65 AA.
XX
AC   ABP07322;
XX
DT   24-JUN-2002  (first entry)
XX
DE   Human ORFX protein sequence SEQ ID NO:14626.
XX
KW   Human; open reading frame; ORFX; gene therapy; cancer; cirrhosis;
KW   hyperproliferative disorder; psoriasis; benign tumour; haemorrhage;
KW   degenerative disorder, osteoarthritis, neurodegenerative disorder;
KW   cardiovascular disease, diabetes mellitus; systemic lupus erythematosus,
KW   hypertension; hypothyroidism, cholesterol ester storage disease,
KW   immune deficiency; immune disorder, infectious disease,
KW   autoimmune disorder; rheumatoid arthritis; autoimmune thyroiditis;
KW   myasthenia gravis.
XX
OS   Homo sapiens.
XX
PN   WO200192523-A2.
XX
PD   06-DEC-2001.
XX
PF   29-MAY-2001; 2001WO-US10836.
XX
PR   30-MAY-2000; 2000US-206132P.
PR   29-AUG-2000; 2000US-228710P.
XX
PA   (CURA-) CURAGEN CORP.
XX
PI   Shimkets RA,  Leach MD;
XX
DR   WPI; 2002-106308/14.
DR   N-PSDB; ABN23074.
XX
PT   Novel human polypeptides and polynucleotides useful for diagnosing,
PT   preventing and treating cardiovascular disease, neurodegenerative,
PT   hyperproliferative disorders and autoimmune disorders -
XX
PS   Disclosure; SEQ ID 14626; 1037pp; English.
XX
CC   The present invention describes substantially purified human proteins
CC   (referred to as open reading frame, ORFX, where X is 1-11491 (see Table 1
CC   in the specification)  ABN15762 to ABN27252 encode the human ORFX
CC   proteins given in ABP00010 to ABP11500. ORFX proteins are useful for
CC   treating or preventing a pathology associated with an ORFX-associated
CC   disorder in humans, and in the manufacture of a medicament for treating a
CC   syndrome associated with ORFX associated disorder. ORFX polynucleotide
CC   sequences can be used in gene therapy. ORFX sequences can be used in the
CC   treatment of cancer, hyperproliferative disorders, cirrhosis of liver,
CC   psoriasis, benign tumours, keloid, degenerative disorders, haemorrhage,
CC   osteoarthritis, neurodegenerative disorders, disorders related to organ
CC   transplantation, cardiovascular diseases, diabetes mellitus, systemic
CC   lupus erythematosus, hypertension, hypothyroidism, cholesterol ester
CC   storage disease, various immune deficiencies and disorders, infectious
CC   diseases, autoimmune disorders such as multiple sclerosis, rheumatoid
CC   arthritis, autoimmune thyroiditis, myasthenia gravis, graft-versus-host
CC   disease and autoimmune inflammatory eye disease. ORFX proteins are also
CC   useful for treating burns, incisions, ulcers, for treating osteoporosis,
CC   bone degenerative disorders, or periodontal disease, and for gut
CC   protection or regeneration and treatment of lung or liver fibrosis,
CC   reperfusion injury in various tissues and conditions resulting from
CC   systemic cytokine damage.
CC   N.B.  The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   65 AA;

  Query Match             100.0%;  Score 38;  DB 23;  Length 65;
  Best Local Similarity   66.7%;  Pred. No. 99;
  Matches    4;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 WXXWHF 6
           |::|||
Db      27 WKSWHF 32


RESULT 14
AAU56391
ID   AAU56391 standard; Protein; 87 AA.
XX
AC   AAU56391;
XX
DT   27-FEB-2002  (first entry)
XX
DE   Propionibacterium acnes immunogenic protein #17287.
XX
KW   SAPHO syndrome, synovitis, acne, pustulosis; hypertosis; osteomyelitis;
KW   uveitis, endophthalmitis, bone, joint, central nervous system; ELISA;
KW   inflammatory lesion; acne vulgaris, enzyme linked immunosorbent assay;
KW   dermatological; osteopathic; neuroprotectant.
XX
OS   Propionibacterium acnes.
XX
PN   WO200181581-A2.
XX
PD   01-NOV-2001.
XX
PF   20-APR-2001, 2001WO-US12865.
XX
PR   21-APR-2000, 2000US-199047P.
PR   02-JUN-2000; 2000US-208841P.
PR   07-JUL-2000; 2000US-216747P.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Skeiky YAW,  Persing DH,  Mitcham JL,  Wang SS,  Bhatia A;
PI   L'maisonneuve J,  Zhang Y,  Jen S,  Carter D;
XX
DR   WPI; 2001-616774/71.
DR   N-PSDB; AAS59575.
XX
PT   Propionibacterium acnes polypeptides and nucleic acids useful for
PT   vaccinating against and diagnosing infections, especially useful for
PT   treating acne vulgaris -
XX
PS   Example 1; SEQ ID No 17586; 1069pp; English.
XX
CC   Sequences AAU89105-AAU68017 represent Propionibacterium acnes immunogenic
CC   polypeptides. The proteins and their associated DNA sequences are used in
CC   the treatment, prevention and diagnosis of medical conditions caused by
CC   P. acnes. The disorders include SAPHO syndrome (synovitis, acne,
CC   pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.
CC   P. acnes is also involved in infections of bone, joints and the central
CC   nervous system, however it is particularly involved in the inflammatory
CC   lesions associated with acne vulgaris. A method for detecting the
CC   presence or absence of P. acnes in a patient comprises contacting a
CC   sample with a binding agent that binds to the proteins of the invention
CC   and determining the amount of bound protein in the sample. The
CC   polypeptides may be used as antigens in the production of antibodies
CC   specific for P. acnes proteins. These antibodies can be used to
```

```
CC   downregulate expression and activity of these polypeptides and
CC   therefore treat or assess infections. The antibodies may also be used as
CC   diagnostic agents for determining for various purposes, for example, by
CC   enzyme linked immunosorbent assay (ELISA).
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   87 AA;

  Query Match             100.0%;  Score 39;  DB 11;  Length 87;
  Best Local Similarity   66.7%;  Pred. No. 1.3e+02;
  Matches    4;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 WXXWHF 6
           |:::||
Db      34 WPPWHF 39


RESULT 15
AAU14777
ID   AAU14777 standard; Protein; 96 AA.
XX
AC   AAU14777;
XX
DT   24-OCT-2001  (first entry)
XX
DE   Novel bone marrow polypeptide #176.
XX
KW   Bone marrow; diagnostic; therapeutic; gene therapy; antigenic;
KW   haematopoiesis; myeloid; lymph cell disorder; tissue regeneration;
KW   wound healing; nutritional supplement; immune disorder;
KW   severe combined immunodeficiency; SCID.
XX
OS   Homo sapiens.
XX
PN   WO200157187-A2.
XX
PD   09-AUG-2001.
XX
PF   05-FEB-2001; 2001WO-US03782.
XX
PR   03-FEB-2000; 2000US-0496914.
PR   20-JUN-2000; 2000US-0598075.
PR   19-JUL-2000; 2000US-0620325.
PR   30-NOV-2000; 2000US-0250683.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Ford JE,  Boyle BJ,  Tang YT,  Liu C,  Asundi V,  Zhou P,  Xue AJ;
PI   Ren F,  Drmanac RT;
XX
DR   WPI; 2001-488875/53.
DR   N-PSDB; AAS23082.
XX
PT   Nucleic acids encoding bone marrow polypeptides, useful in diagnostic
PT   and gene therapy -
XX
PS   Claim 10; Page 137; 392pp; English.
XX
CC   AAU14602-AAU14794 represent novel bone marrow polypeptides of the
CC   invention. The proteins and corresponding coding sequences may be used
CC   in the prevention, diagnosis and treatment of diseases associated with
CC   inappropriate bone marrow polypeptide expression. For example, to treat
CC   disorders associated with decreased expression by rectifying mutations
CC   or deletions in a patient's genome that affect the activity of the
CC   polypeptides by expressing inactive proteins or to supplement the
CC   patient's own production of the polypeptide. Additionally, the nucleic
CC   acids may be used to produce the polypeptides, by inserting the nucleic
CC   acids into a host cell and culturing the cell to express the protein.
CC   The nucleic acid and its complementary sequences may also be used as DNA
CC   probes in diagnostic assays to detect and quantitate the presence of
CC   similar nucleic acid sequences in samples, and therefore which patients
CC   may be in need of restorative therapy. The proteins may also be used as
CC   antigens in the production of antibodies against bone marrow proteins
CC   and in assays to identify, quantitate, if their expression and activity.
CC   The anti bone marrow protein antibodies and antagonists may also be used
CC   to down regulate expression and activity. The antibodies may also be used
CC   as diagnostic agents for detecting the presence of the protein in samples
CC   (e.g. by enzyme linked immunosorbant assay (ELISA)). The proteins
CC   may be used to regulate haematopoietic activity, and consequently in the
CC   treatment of myeloid or lymph cell disorders, in tissue regeneration,
CC   such as wound healing, as a nutritional supplement, and in treatment of
CC   immune disorders such as severe combined immunodeficiency (SCID).
XX
SQ   Sequence   96 AA;

  Query Match             100.0%;  Score 39;  DB 11;  Length 96;
  Best Local Similarity   66.7%;  Pred. No. 1.4e+02;
  Matches    4;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 WXXWHF 6
           |:::||
Db      67 WLIWHF 72


Search completed: February 20, 2003, 10:47:37
Job time : 30.3333 secs
```

GenCore version 5.1.3

OM protein - protein search, using sw model

Run on: February 21, 2003, 11:24:12 ; Search Time 11 Seconds
(without alignments)
48.067 Million cell updates/sec

Title: US-09-912-414-11
Perfect score: 6
Sequence: 1 WXXWHF 6

Scoring table: OLIGOPX
Gapop 60.0 , Gapext 60.0

Searched: 283224 seqs, 96134422 residues

Word size : 0

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : PIR_73:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	6	100.0	230	2	D70847	hypothetical prote
2	6	100.0	241	2	A71020	hypothetical prote
3	6	100.0	260	2	B71353	probable D,D carbo
4	6	100.0	293	2	A83299	hypothetical prote
5	6	100.0	358	1	S43876	site-specific DNA
6	6	100.0	358	2	A87296	modification methy
7	6	100.0	386	2	A97456	adenine-specific m
8	6	100.0	386	2	AD2674	adenine DNA methyl
9	6	100.0	394	2	D90105	putative SAR DNA-b
10	6	100.0	403	2	AF3432	site-specific DNA-
11	6	100.0	423	2	AH1407	PTS system galacti
12	6	100.0	423	2	AH1783	PTS system galacti
13	6	100.0	424	2	JC5891	omega 6 desaturase
14	6	100.0	424	2	T07742	omega-6 desaturase
15	6	100.0	439	2	T01807	hypothetical prote
16	6	100.0	441	2	T08136	probable omega-6 d
17	6	100.0	447	2	T17299	hypothetical prote
18	6	100.0	448	2	D85362	hypothetical prote
19	6	100.0	452	2	H91172	probable phosphotr
20	6	100.0	452	2	AH0964	probable PTS syste
21	6	100.0	462	2	H86018	hypothetical prote
22	6	100.0	472	2	AG2417	hypothetical prote
23	6	100.0	475	2	B84171	hypothetical prote
24	6	100.0	487	2	T32941	hypothetical prote
25	6	100.0	489	2	B85540	hypothetical prote
26	6	100.0	492	2	AH2079	alpha-amylase [imp
27	6	100.0	497	2	T40586	nucleolar protein
28	6	100.0	504	2	S48550	hypothetical prote
29	6	100.0	508	2	G90372	4-hydroxyphenylace
30	6	100.0	508	2	T50180	nucleolar protein
31	6	100.0	511	2	S59322	nucleolar protein
32	6	100.0	522	2	D96602	nucleolar protein
33	6	100.0	550	2	T06379	SAR DNA-binding pr
34	6	100.0	560	2	T06377	SAR DNA-binding pr
35	6	100.0	640	2	T28621	Y4cD protein Rhi
36	6	100.0	718	2	T29449	hypothetical prote
37	6	100.0	774	2	A81979	hypothetical prote
38	5	83.3	40	2	F81511	hypothetical prote
39	5	83.3	41	2	D82458	hypothetical prote
40	5	83.3	42	2	S02108	cytochrome c oxida
41	5	83.3	47	2	T04365	ethylene responsiv
42	5	83.3	50	2	A95009	hypothetical prote
43	5	83.3	52	2	F97980	hypothetical prote
44	5	83.3	53	2	F80871	hypothetical prote
45	5	83.3	54	2	G92682	hypothetical prote

ALIGNMENTS

RESULT 1
D70847
hypothetical protein Rv0059 - Mycobacterium tuberculosis (strain H37RV)
C;Species: Mycobacterium tuberculosis
C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C;Accession: D70847
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A;Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A;Reference number: A70500; MUID:98209987; PMID:9634230
A;Accession: D70847
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-230 <JDL>
A;Cross-references: GB:AL021408; GB:AL123456; NID:g2621514; PIDN:CAA15240.1; PID:g2800710
A;Experimental source: strain H37Rv
C;Genetics:
A;Gene: Rv0059
C;Superfamily: Mycobacterium tuberculosis hypothetical protein Rv0059

Query Match 100.0%; Score 6; DB 2; Length 230;
Best Local Similarity 66.7%; Pred. No. 46;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 WXXWHF 6
|::|||
Db 26 WIVWHF 31

RESULT 2
A71020
hypothetical protein PH1454 - Pyrococcus horikoshii
C;Species: Pyrococcus horikoshii
C;Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 20-Jun-2000
C;Accession: A71020
R;Kawarabayasi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; Sekine
M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi,
DNA Res. 5, 55-76, 1998
A;Title: Complete sequence and gene organization of the genome of a hyper-thermophilic ar
A;Reference number: A71000; MUID:98344137; PMID:9679194
A;Accession: A71020
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-241 <KAW>
A;Cross-references: GB:AP000006; NID:g3256133; PIDN:BAA29154.1; PID:g3258878
A;Experimental source: strain OT3
A;Note: this accession replaces an interim accession for a sequence replaced by GenBank
C;Genetics:
A;Gene: PH1454

```
  Query Match             100.0%;  Score 6;  DB 2;  Length 201;
  Best Local Similarity   66.7%;  Pred. No. 99;
  Matches    4;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 WXXWHF 6
           |::|||
Db     193 WQTWHF 198


RESULT 3
B71353
probable D,D-dipeptidase - syphilis spirochete
C;Species: Treponema pallidum subsp. pallidum (syphilis spirochete)
C;Date: 24-Jul-1998 #sequence_revision 24-Jul-1998 #text_change 05-Nov-1999
C;Accession: B71353
R;Fraser, C.M.; Norris, S.J.; Weinstock, G.M.; White, O.; Sutton, G.G.; Dodson, R.; Gw
rson, J.; Khalak, H.; Richardson, D.; Howell, J.K.; Chidambaram, M.; Utterback, T.; McD
they, L.; Weidman, J.; Smith, H.O.; Venter, J.C.
Science 281, 375-388, 1998
A;Title: Complete genome sequence of Treponema pallidum, the syphilis spirochete.
A;Reference number: A71250; MUID:98332770; PMID:9665876
A;Accession: B71353
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-260 <COL>
A;Cross-references: GB:AE000993; GB:AE000993; NID:g3322870; PIDN:AAC65219.1; PID:g332
A;Experimental source: strain Nichols
C;Genetics:
A;Gene: TP0221

  Query Match             100.0%;  Score 6;  DB 2;  Length 225;
  Best Local Similarity   66.7%;  Pred. No. 1e+02;
  Matches    4;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 WXXWHF 6
           ::
Db     220 WEPWHF 225


RESULT 4
A83299
hypothetical protein PA2778 [imported] - Pseudomonas aeruginosa (strain PAO1)
C;Species: Pseudomonas aeruginosa
C;Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C;Accession: A83299
R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrener, P.; Hickey, M.J.; Br
adman, S.; Yuan, Y.; Erdy, J.; Chilton, C.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A;Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
A;Reference number: A75700; MUID:20405561; PMID:10984043
A;Accession: A83299
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-293 <STO>
A;Cross-references: GB:AE004778; GB:AE004141; NID:g9951141; PID:AAG04641.1; GSPDB:GN0
A;Experimental source: strain PAO1
C;Genetics:
A;Gene: PA2778

  Query Match             100.0%;  Score 6;  DB 2;  Length 293;
  Best Local Similarity   66.7%;  Pred. No. 1e+02;
  Matches    4;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 WXXWHF 6
           |::|||
Db     101 WPRWHF 106


RESULT 5
S43876
site-specific DNA methyltransferase (EC 2.1.1.-) - Caulobacter crescentus
C;Species: Caulobacter crescentus
C;Date: 16-Sep-1994 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C;Accession: S43876
R;Zweiger, G.; Marczynski, G.; Shapiro, L.
J. Mol. Biol. 235, 472-485, 1994
A;Title: A Caulobacter DNA methyltransferase that functions only in the predivisional cel
A;Reference number: S43876; MUID:94116303; PMID:8289276
A;Accession: S43876
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-358 <ZWE>
A;Cross-references: EMBL:U01032; NID:g393011; PIDN:AAA19913.1; PID:g393012
C;Superfamily: site specific DNA-methyltransferase (adenine-specific); HpaI
C;Keywords: methyltransferase; restriction-modification system

  Query Match             100.0%;  Score 6;  DB 1;  Length 358;
  Best Local Similarity   66.7%;  Pred. No. 1.3e+02;
  Matches    4;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 WXXWHF 6
           |::|||
Db     332 WTYWHF 337


RESULT 6
A87296
adenine specific methylase CcrM [imported] - Caulobacter crescentus
C;Species: Caulobacter crescentus
C;Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 17-May-2002
C;Accession: A87296
R;Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.
F.; Lanh, M.J.; Ketchy, R.T.; Smit, J.; Dimitrov, G.I.; Ward, N.L.; Hoft, D.H.; Kolona
s, J.; Brinkac, M.; White, O.; Salzberg, S.L.; Chapin, T.; Venter, J.C.; Fraser, C.M.
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A;Title: Complete Genome Sequence of Caulobacter crescentus.
A;Reference number: A87249; MUID:21173608; PMID:11259647
A;Accession: A87296
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-358 <STO>
A;Cross-references: GB:AE006673; NID:g13421033; PIDN:AAK22965.1; GSPDB:GN00146
C;Genetics:
A;Gene: CC0378
C;Superfamily: site specific DNA methyltransferase (adenine specific); HpaI

  Query Match             100.0%;  Score 6;  DB 2;  Length 358;
  Best Local Similarity   66.7%;  Pred. No. 1.3e+02;
  Matches    4;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 WXXWHF 6
           |::|||
Db     332 WTYWHF 337


RESULT 7
A97456
adenine specific methyltransferase CMT1 (= Rhizobium methylase CMT1) [imported] - Agrob
C;Species: Agrobacterium tumefaciens
C;Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 17-May-2002
C;Accession: A97456
R;Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman,
A.; Liu, Y.; Wallam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, T.;
Science 294, 2323-2328, 2001
A;Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tume
A;Reference number: A97359; PMID:11743194
A;Accession: A97456
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-386 <KUR>
A;Cross-references: GB:AE007869; PIDN:AAK86602.1; PID:g15156772; GSPDB:GN00169
C;Genetics:
A;Gene: AGR_C_1463
A;Map position: circular chromosome
```

C;Superfamily: site-specific DNA-methyltransferase (adenine-specific) HpaI

Query Match 100.0%; Score 6; DB 2; Length 386;
Best Local Similarity 66.7%; Pred. No. 1.4e+02;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 WXXWHF 6
|::|||
Db 357 WTFWHF 362

RESULT 8
AD2674
adenine DNA methyltransferase [imported] Agrobacterium tumefaciens (strain C58, Dupont
C;Species: Agrobacterium tumefaciens
C;Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 17-May-2002
C;Accession: AD2674
R;Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Wag, L
erage, G.; Gillet, W.; Grant, C.; Guenthner, D.; Kutyavin, T.; Levy, R.; Li, M.; McClell
; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A;Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon Kamm,
ster, E.W.
A;Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58
A;Reference number: AB2577; PMID:11743193
A;Accession: AD2674
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-386 <KUR>
A;Cross-references: GB:AE009669; PIDN:AAL41810.1; PID:g17733167; GSPDB:GN00186
A;Experimental source: strain C58 (Dupont)
C;Genetics:
A;Gene: Atu0794
A;Map position: circular chromosome
C;Superfamily: site-specific DNA-methyltransferase (adenine-specific) HpaI

Query Match 100.0%; Score 6; DB 2; Length 386;
Best Local Similarity 66.7%; Pred. No. 1.4e+02;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 WXXWHF 6
|::|||
Db 357 WTFWHF 362

RESULT 9
D90105
putative SAP DNA binding protein 1 [imported] - Guillardia theta nucleomorph
C;Species: nucleomorph Guillardia theta
A;Note: a nucleomorph is the vestigial nucleus of a eukaryotic endosymbiont
C;Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 24-May-2001
C;Accession: D90105
R;Douglas, S.; Zauner, S.; Fraunholz, M.; Beaton, M.; Penny, S.; Deng, L.T.; Wu, X.; Fe
Nature 410, 1091-1096, 2001
A;Title: The highly reduced genome of an enslaved algal nucleus.
A;Reference number: A90080; MUID:11323671; PMID:11323671
A;Accession: D90105
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-394 <DOU>
A;Cross-references: GB:AF041553; NID:g7529673; PIDN:AAF61781.1; GSPDB:GN00151
C;Genetics:
A;Map position: 2
A;Genome: nucleomorph
C;Keywords: nucleomorph

Query Match 100.0%; Score 6; DB 2; Length 394;
Best Local Similarity 66.7%; Pred. No. 1.4e+02;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 WXXWHF 6
|::|||
Db 169 WYSWHF 174

RESULT 10
AF3432
site-specific DNA methyltransferase (adenine-specific) (EC 2.1.1.72) [imported] - Brucell
C;Species: Brucella melitensis
C;Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 17-May-2002
C;Accession: AF3432
R;DelVecchio, V.G.; Kapatral, V.; Redkar, R.J.; Patra, G.; Mujer, C.; Los, T.; Ivanova, N
.; Mazur, M.; Goltsman, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Joseph
Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002
A;Title: The genome sequence of the facultative intracellular pathogen Brucella melitensi
A;Reference number: AB3252; PMID:11756688
A;Accession: AF3432
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-403 <KUR>
A;Cross-references: GB:AE008917; PIDN:AAL52635.1; PID:g17983446; GSPDB:GN00196
A;Experimental source: strain 16M
C;Genetics:
A;Gene: BMEI1444
A;Map position: I
C;Superfamily: site-specific DNA methyltransferase (adenine-specific) HpaI
C;Keywords: methyltransferase; S-adenosylmethionine

Query Match 100.0%; Score 6; DB 2; Length 403;
Best Local Similarity 66.7%; Pred. No. 1.5e+02;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 WXXWHF 6
|::|||
Db 373 WTFWHF 378

RESULT 11
AH1407
PTS system galactitol specific enzyme IIC component homolog lmo2665 [imported] - Listeria
C;Species: Listeria monocytogenes
C;Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 14-Dec-2001
C;Accession: AH1407
R;Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloecker,
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.;
D.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A;Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Mar
ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,
A;Title: Comparative genomics of Listeria species.
A;Reference number: AB1077; MUID:21647274; PMID:11677669
A;Accession: AH1407
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-423 <GLA>
A;Cross-references: GB:NC_003210; PIDN:AAM00078.1; PID:g16411214; GSPDB:GN00177
A;Experimental source: strain EGD-e
C;Genetics:
A;Gene: lmo2665
C;Superfamily: phosphotransferase enzyme II galactitol specific

Query Match 100.0%; Score 6; DB 2; Length 423;
Best Local Similarity 66.7%; Pred. No. 1.5e+02;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 WXXWHF 6
|::|||
Db 126 WNYWHF 131

RESULT 12
AH1783
PTS system galactitol-specific enzyme IIC component homolog lin2814 [imported] - Listeria
C;Species: Listeria innocua
C;Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 14-Dec-2001
C;Accession: AH1783

R;Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloecker
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.
D.; Jones, L.M.; Karst, U.
Science 294, 849 852, 2001
A;Authors: Pfeft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maltsburham, A.; Ma
ck, C.; Schluter, T.; Simoud, N.; Tiollais, A.; Vazquez Boland, J.A.; Voss, H.; Wolland,
A;Title: Comparative genomics of listeria species.
A;Reference number: AB... ?; MUID:...; PMID:...
A;Accession: AH1783
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-423 <GLA>
A;Cross-references: GB:AL592022; PIDN:CAC99040.1; PID:g16411281; GSPDB:GN00179
A;Experimental source: strain Clip11262
C;Genetics:
A;Gene: lin2814
C;Superfamily: phosphotransferase enzyme II galactitol specific

  Query Match 100.0%; Score 6; DB 2; Length 423;
  Best Local Similarity ...%; Pred. No. 1.5e+02;
  Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 WXXWHF 6
       |:|||
Db 126 WNYWHF 131

RESULT 13
JC5891
omega-6 desaturase (EC 1.14.99.-) precursor, chloroplast - Chlamydomonas reinhardtii
C;Species: Chlamydomonas reinhardtii
C;Date: 18-Mar-1998 #sequence_revision 18-Mar-1998 #text_change 30-Jun-2000
C;Accession: JC5891
R;Sato, N.; Fujiwara, S.; Kawaguchi, A.; Tsuzuki, M.
J. Biochem. 122, 1224 1232, 1997
A;Title: Cloning of a gene for chloroplast omega-6 desaturase of a green alga, Chlamydom
A;Reference number: JC5891; MUID:98158334; PMID:9498569
A;Accession: JC5891
A;Status: nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 1-424 <SAT>
A;Cross-references: DBJ:AB... ...; NID:g...; PIDN:BAA...; PID:g...
C;Comment: This enzyme catalyzes the desaturation of monoenoic to dienoic acids.
C;Genetics:
A;Gene: des6
C;Superfamily: omega 3 fatty acid desaturase
C;Keywords: chloroplast, oxidoreductase
F;1-40/Domain: transit peptide (chloroplast) #status predicted <TPD>

  Query Match 100.0%; Score 6; DB 2; Length 424;
  Best Local Similarity 66.7%; Pred. No. 1.5e+02;
  Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 WXXWHF 6
       |:|||
Db 234 WLVWHF 239

RESULT 14
T07742
omega-6 desaturase, chloroplast - soybean
C;Species: Glycine max (soybean)
C;Date: 14-May-1999 #sequence_revision 14-May-1999 #text_change 08-Oct-1999
C;Accession: T07742
R;Hitz, W.D.; Carlson, T.J.; Booth, J.R.; Kinney, A.J., Shoaca, K.L., Yadav, N.S.
Plant Physiol. 105, 635-641, 1994
A;Title: Cloning of a higher plant plastid omega-6 fatty acid desaturase cDNA and its ex
A;Reference number: Z16109; MUID:94345008; PMID:8066133
A;Accession: T07742
A;Status: preliminary; translated from GP/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-424 <HIT>
A;Cross-references: EMBL:L29215; NID:g459961; PIDN:AAA50168.1; PID:g459962
A;Experimental source: seed
C;Genetics:
A;Genome: nuclear
C;Superfamily: omega 3 fatty acid desaturase
C;Keywords: chloroplast

  Query Match 100.0%; Score 6; DB 2; Length 424;
  Best Local Similarity ...%; Pred. No. 1.5e+02;
  Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 WXXWHF 6
       |:|||
Db 351 WLMWHF 356

RESULT 15
T01807
hypothetical protein A_TM021B04.12 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 26-Feb-1999 #sequence_revision 26-Feb-1999 #text_change 02-Jun-2000
C;Accession: T01807
R;Dante, M.; Wamsley, P.; Gibson, A.
submitted to the EMBL Data Library, June 1997
A;Description: The sequence of A. thaliana TM021B04.
A;Reference number: Z14440
A;Accession: T01807
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-439 <DAN>
A;Cross-references: EMBL:AF007271; NID:g2191191; PIDN:g2191199; GSPDB:GN00093; ATH1:A_TM02
C;Genetics:
A;Gene: AT3F:A_TM021B04.12
A;Map position: 5
A;Introns: 24/3; 46/3; 109/3; 216/3; 282/3; 388/3
C;Superfamily: garden pea DAP DNA binding protein

  Query Match 100.0%; Score 6; DB 2; Length 439;
  Best Local Similarity 66.7%; Pred. No. 1.6e+02;
  Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 WXXWHF 6
       |:|||
Db 184 WFCWHF 189

Search completed: February 20, 2003, 11:29:40
Job time : 12 secs

GenCore version 5.1.3

OM protein - protein search, using sw model

Run on: February 20, 2003, 11:20:17 ; Search time 11.3333 Seconds
(without alignments)
21.958 Million cell updates/sec

Title: US-09-912-414-11
Perfect score: 6
Sequence: 1 WXXWHF 6

Scoring table: OLIGODX
Gapop 60.0 , Gapext 60.0

Searched: 112892 seqs, 41476328 residues

Word size : 0

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No	Score	% Query Match	Length	DB	ID	Description
1	6	100.0	358	1	MTC1_CAUCR	Q45971 caulobacter
2	6	100.0	376	1	MTS1_RHIME	Q30569 rhizobium m
3	6	100.0	377	1	MTR1_BPUAR	Q30570 brucella ab
4	6	100.0	424	1	FD6C_SOYBN	P48628 glycine max
5	6	100.0	439	1	Y412_ARATH	Q94658 arabidopsis
6	6	100.0	443	1	FD6C_BRANA	P48627 brassica na
7	6	100.0	448	1	FD6C_ARATH	P46312 arabidopsis
8	6	100.0	479	1	CATA_PSEPU	Q59714 pseudomonas
9	6	100.0	504	1	SIK1_YEAST	Q12460 saccharomyc
10	6	100.0	511	1	NOP5_YEAST	Q12499 saccharomyc
11	6	100.0	529	1	NOP5_HUMAN	Q9y2x3 homo sapien
12	6	100.0	534	1	NOP5_RAT	Q9qz86 rattus norv
13	6	100.0	640	1	Y4CD_RHISN	P55386 rhizobium s
14	6	100.0	718	1	PLSB_CAEEL	Q22049 caenorhabdi
15	5	83.3	82	1	TXA1_ACTPQ	Q9rjq2 actinia equ
16	5	83.3	85	1	VG73_BPML5	Q05288 mycobacteri
17	5	83.3	95	1	ESA6_MYCLE	Q50206 mycobacteri
18	5	83.3	97	1	YC49_CYACA	Q9tm18 cyanidium c
19	5	83.3	101	1	ATPK_YEAST	Q06405 saccharomyc
20	5	83.3	103	1	ES6D_MYCTU	O05440 mycobacteri
21	5	83.3	104	1	Y49L_SYNY3	Q55720 synechocyst
22	5	83.3	104	1	YJA7_YEAST	P47080 saccharomyc
23	5	83.3	106	1	NIGM_HUMAN	Q95178 homo sapien
24	5	83.3	108	1	NIGM_BOVIN	Q02374 bos taurus
25	5	83.3	111	1	VPX_HV2KR	Q74122 human immun
26	5	83.3	111	1	YFBW_SALTI	P81891 salmonella
27	5	83.3	111	1	YFBW_SALTY	O52328 salmonella
28	5	83.3	112	1	TX1A_AGEAP	P15969 agelenopsis
29	5	83.3	112	1	VPX_HV2CA	P24110 human immun
30	5	83.3	112	1	VPX_HV2D1	P17760 human immun
31	5	83.3	112	1	VPX_HV2G1	P18045 human immun
32	5	83.3	112	1	VPX_HV2N2	P05919 human immun
33	5	83.3	112	1	VPX_HV2RO	P06939 human immun
34	5	83.3	112	1	VPX_HV2SB	P12454 human immun
35	5	83.3	112	1	VPX_HV2ST	P20881 human immun
36	5	83.3	112	1	VPX_SIVM1	P05917 simian immu
37	5	83.3	112	1	VPX_SIVMK	P05916 simian immu
38	5	83.3	112	1	VPX_SIVML	P11266 simian immu
39	5	83.3	112	1	VPX_SIVS4	P12514 simian immu
40	5	83.3	112	1	VPX_SIVSP	P19508 simian immu
41	5	83.3	113	1	VPX_HV1BE	P19939 human immun
42	5	83.3	118	1	PA2I_NAJME	P00599 naja melano
43	5	83.3	118	1	PA21_NAJMO	P00602 naja mossam
44	5	83.3	118	1	PA22_NAJMO	P00603 naja mossam
45	5	83.3	118	1	PA23_NAJMO	P00604 naja mossam

ALIGNMENTS

RESULT 1
MTC1_CAUCR
ID MTC1_CAUCR STANDARD; PRT; 358 AA.
AC Q45971;
DT 30-MAY-2000 (Rel. 39, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Modification methylase CcrmI (EC 2.1.1.72) (Adenine-specific
DE methyltransferase CcrmI) (M.CcrmI).
GN CCRMIM OR CCRM OR CC0378.
OS Caulobacter crescentus.
OC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;
OC Caulobacter.
OX NCBI_TaxID=155892;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CB15N / NA1000;
RX MEDLINE=94118303; PubMed=8289276;
RA Zweiger G., Marczynski G., Shapiro L.;
RT "A Caulobacter DNA methyltransferase that functions only in the
RT predivisional cell.";
RL J. Mol. Biol. 235:472-485(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 19089 / CB15;
RX MEDLINE 21173698, PubMed=11259647;
RA Nierman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
RA Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
RA Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
RA DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,
RA Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,
RA Utterback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,
RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
RT "Complete genome sequence of Caulobacter crescentus.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
CC -!- FUNCTION: THIS METHYLASE RECOGNIZES THE DOUBLE STRANDED SEQUENCE
CC GANTC AND CAUSES SPECIFIC METHYLATION ON A 2 ON BOTH STRANDS.
CC CCRM-MEDIATED METHYLATION HAS IMPORTANT CELLULAR FUNCTIONS.
CC APPEARS TO CONTRIBUTE TO THE ACCURATE CELL CYCLE CONTROL OF DNA
CC REPLICATION AND CELLULAR MORPHOLOGY.
CC -!- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + DNA adenine = S-
CC adenosyl-L-homocysteine + DNA 6-methylaminopurine.
CC ----------
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC ----------
DR EMBL; U01032; AAA18913.1; -.
DR EMBL; AE005711; AAK22365.1; -.
DR HSSP; P11409; 1BOO.
DR REBASE; 2539; M.CcrMI.
DR TIGR; CC0378; -.

```
DR   InterPro; IPR001091; CN4_Metransf.
DR   InterPro; IPR002295; D21N6_mtfrase.
DR   InterPro; IPR002941; N6/N4_Mtase.
DR   InterPro; IPR002052; N6_Mtase.
DR   InterPro; IPR000051; SAM_bind.
DR   Pfam; PF01555, N6_N4_Mtase, 1.
DR   PRINTS; PR00506, D21N6MTFRASE.
DR   PRINTS; PR00508, S21N4MTFRASE.
DR   PROSITE; PS01124, N6_MTASE, 1.
KW   Transferase, Methyltransferase, DNA replication, Complete proteome.
FT   CONFLICT     242    242       Y -> D (IN REF. 3).
SQ   SEQUENCE   358 AA,  39??? MW,  ??F4?????C4????4 ?????,

  Query Match:          100.0%;  Score 6;  DB 1;  Length 358;
  Best Local Similarity   66.7%;  Pred. No. 41;
  Matches    4;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 WXXWHF 6
             :: :
Db     332 WTYWHF 337


RESULT 2
MTS1_RHIME
ID   MTS1_RHIME     STANDARD;      PRT;   376 AA.
AC   O30569;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Modification methylase SmeIP (EC 2.1.1.72) (Adenine-specific
DE   methyltransferase SmeIP) (M.SmeI) (M.CcrMI).
GN   SMAIM OR CCRM OR RA0026 OR SMC00??.
OS   Rhizobium meliloti (Sinorhizobium meliloti).
OC   Bacteria; Proteobacteria, alpha subdivision, Rhizobiaceae group,
OC   Rhizobiaceae; Sinorhizobium.
OX   NCBI_TaxID 382,
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=1021;
RX   MEDLINE 97440133, PubMed 9294447,
RA   Wright R., Stephens C., Shapiro L.;
RT   "The CcrM DNA methyltransferase is widespread in the alpha subdivision
RT   of proteobacteria, and its essential functions are conserved in
RT   Rhizobium meliloti and Caulobacter crescentus.",
RL   J. Bacteriol. 179:5869-5877(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=1021;
RX   MEDLINE 21368147, PubMed=11404447;
RA   Capela D., Barloy-Hubler F., Gouzy J., Bothe G., Ampe F., Batut J.,
RA   Boistard P., Becker A., Boutry M., Cadieu E., Dreano S., Gloux S.,
RA   Godrie T., Goffeau A., Kahn D., Kiss E., Lelaure V., Masuy D.,
RA   Pohl T., Portetelle D., Puhler A., Purnelle B., Ramsperger U.,
RA   Renard C., Thebault P., Vandenbol M., Weidner S., Galibert F.,
RT   "Analysis of the chromosome sequence of the legume symbiont
RT   Sinorhizobium meliloti strain 1021.",
RL   Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882(2001).
CC   -!- FUNCTION: THIS METHYLASE RECOGNIZES THE DOUBLE STRANDED SEQUENCE
CC       GANTC AND CAUSES SPECIFIC METHYLATION ON A 2 ON BOTH STRANDS.
CC       CCRM-MEDIATED METHYLATION HAS IMPORTANT CELLULAR FUNCTIONS.
CC       APPEARS TO CONTRIBUTE TO THE ACCURATE CELL CYCLE CONTROL OF DNA
CC       REPLICATION AND CELLULAR MORPHOLOGY.
CC   -!- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + DNA adenine = S-
CC       adenosyl-L-homocysteine + DNA 6-methylaminopurine.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AF011894; AAB71350.1; -.
DR   EMBL; AL591795; CAC45498.1; -.
DR   REBASE; 3264; M.SmeIP.
DR   InterPro; IPR001091; CN4_Metransf.
DR   InterPro; IPR002295; D21N6_mtfrase.
DR   InterPro; IPR002941; N6/N4_Mtase.
DR   InterPro; IPR002052; N6_Mtase.
DR   InterPro; IPR000051; SAM_bind.
DR   Pfam; PF01555, N6_N4_Mtase; 1.
DR   PRINTS; PR00506, D21N6MTFRASE.
DR   PRINTS; PR00508, S21N4MTFRASE.
DR   PROSITE; PS01124, N6_MTASE, 1.
KW   Transferase; Methyltransferase, DNA replication, Complete proteome.
FT   CONFLICT     136    141       NPMFNFK -> QFDADLQ (IN REF. 1).
FT   CONFLICT     157    157       P -> A (IN REF. 1).
SQ   SEQUENCE   376 AA,  41446 MW,  ????47F8?C?? A?79C4;

  Query Match:          100.0%;  Score 6;  DB 1;  Length 376;
  Best Local Similarity   66.7%;  Pred. No. 42;
  Matches    4;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 WXXWHF 6
             :: |||
Db     347 WTFWHF 352


RESULT 3
MTB1_BRUAB
ID   MTB1_BRUAB     STANDARD;      PRT;   377 AA.
AC   O30570;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Modification methylase BabI (EC 2.1.1.72) (Adenine-specific
DE   methyltransferase BabI) (M.BabI) (M.CcrMI).
GN   BABIM OR CCRM.
OS   Brucella abortus.
OC   Bacteria; Proteobacteria, alpha subdivision, Rhizobiaceae group;
OC   Brucellaceae; Brucella.
OX   NCBI_TaxID 235;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S2308;
RX   MEDLINE 97440133, PubMed 9294447;
RA   Wright R., Stephens C., Shapiro L.;
RT   "The CcrM DNA methyltransferase is widespread in the alpha subdivision
RT   of proteobacteria, and its essential functions are conserved in
RT   Rhizobium meliloti and Caulobacter crescentus.";
RL   J. Bacteriol. 179:5869-5877(1997).
CC   -!- FUNCTION: THIS METHYLASE RECOGNIZES THE DOUBLE-STRANDED SEQUENCE
CC       GANTC AND CAUSES SPECIFIC METHYLATION ON A 2 ON BOTH STRANDS.
CC       CCRM-MEDIATED METHYLATION HAS IMPORTANT CELLULAR FUNCTIONS.
CC       APPEARS TO CONTRIBUTE TO THE ACCURATE CELL CYCLE CONTROL OF DNA
CC       REPLICATION AND CELLULAR MORPHOLOGY.
CC   -!- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + DNA adenine = S-
CC       adenosyl-L-homocysteine + DNA 6-methylaminopurine.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AF011895; AAB71351.1; -.
DR   REBASE; 3265; M.BabI.
DR   InterPro; IPR001091; CN4_Metransf.
DR   InterPro; IPR002295; D21N6_mtfrase.
DR   InterPro; IPR002941; N6/N4_Mtase.
DR   InterPro; IPR002052; N6_Mtase.
DR   InterPro; IPR000051; SAM_bind.
DR   Pfam; PF01555; N6_N4_Mtase; 1.
```

```
DR   PRINTS; PR00506; D21N6MTFRASE.
DR   PRINTS; PR00508; S21N4MTFRASE.
DR   PROSITE; PS00092; N6_MTASE; 1
KW   Transferase; Methyltransferase; DNA replication.
SQ   SEQUENCE   377 AA;  42202 MW;  [illegible] CRC64;

  Query Match           100.0%;  Score 6;  DB 1;  Length 377;
  Best Local Similarity   66.7%;  Pred. No. 42;
  Matches    4;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 WXXWHF 6
          |::|||
Db    347 WTFWHF 352


RESULT 4
FD6C_SOYBN
ID   FD6C_SOYBN     STANDARD;      PRT;   424 AA.
AC   P48628;
DT   01-FEB-1996 (Rel. 33, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Omega-6 fatty acid desaturase, chloroplast precursor (EC 1.14.99.-).
OS   Glycine max (Soybean)
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.
OX   NCBI_TaxID=3847;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Seed;
RX   MEDLINE=94345008; PubMed=8066133;
RA   Hitz W.D., Carlson T.J., Booth J.R. Jr., Kinney A.J., Stecca K.L.,
RA   Yadav N.S.;
RT   "Cloning of a higher-plant plastid omega-6 fatty acid desaturase cDNA
RT   and its expression in a cyanobacterium.";
RL   Plant Physiol. 105:635-641(1994).
CC   -!- FUNCTION: CHLOROPLAST OMEGA-6 FATTY ACID DESATURASE INTRODUCES
CC       THE SECOND DOUBLEBOND IN THE BIOSYNTHESIS OF 16.3 AND 18.3 FATTY
CC       ACIDS, IMPORTANT CONSTITUENTS OF PLANT MEMBRANES. IT IS THOUGHT
CC       TO USE FERREDOXIN AS AN ELECTRON DONOR AND TO ACT ON FATTY ACIDS
CC       ESTERIFIED TO GALACTOLIPIDS, SULFOLIPIDS AND PHOSPHATIDYLGLYCEROL.
CC   -!- PATHWAY: Polyunsaturated fatty acid biosynthesis.
CC   -!- SUBCELLULAR LOCATION: CHLOROPLAST, MEMBRANE-BOUND (PROBABLE)
CC   -!- DOMAIN: THE HISTIDINE BOX DOMAINS MAY CONTAIN THE ACTIVE SITE
CC       AND/ OR BE INVOLVED IN METAL ION BINDING.
CC   -!- SIMILARITY: BELONGS TO THE FATTY ACID DESATURASE FAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; L29215; AAA50158.1; -.
DR   InterPro; IPR001225; FA_desaturase.
DR   Pfam; PF00487; FA_desaturase; 1.
DR   ProDom; PD001081; FA_desaturase; 1.
KW   Oxidoreductase; Fatty acid biosynthesis; Chloroplast; Membrane;
KW   Transit peptide
FT   TRANSIT       1     63       CHLOROPLAST (BY SIMILARITY).
FT   CHAIN        64    424       OMEGA-6 FATTY ACID DESATURASE.
FT   DOMAIN      165    169       HISTIDINE BOX-1.
FT   DOMAIN      201    205       HISTIDINE BOX-2.
FT   DOMAIN      361    365       HISTIDINE BOX-3.
SQ   SEQUENCE   424 AA;  49641 MW;  4F3DF52D4B1A2009 CRC64;

  Query Match           100.0%;  Score 6;  DB 1;  Length 424;
  Best Local Similarity   66.7%;  Pred. No. 46;
  Matches    4;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 WXXWHF 6
          |::|||
Db    251 WLMWHF 256


RESULT 5
Y412_ARATH
ID   Y412_ARATH     STANDARD;      PRT;   439 AA.
AC   Q04658;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Hypothetical protein At5g27120
GN   AT5G27120 OR TM021B04.12 OR T21B4.30.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Columbia;
RX   MEDLINE=21016721; PubMed=11130714;
RA   Tabata S., Kaneko T., Nakamura Y., Kotani H., Kato T., Asamizu E.,
RA   Miyajima N., Sasamoto S., Kimura T., Hosouchi T., Kawashima K.,
RA   Kohara M., Matsumoto M., Matsuno A., Muraki A., Nakayama S.,
RA   Nakazaki N., Naruo K., Okumura S., Shinpo S., Takeuchi C., Wada T.,
RA   Watanabe A., Yamada M., Yasuda M., Sato S., de la Bastide M.,
RA   Huang E., Spiegel L., Gnoj L., O'Shaughnessy A., Preston R.,
RA   Habermann K., Murray J., Johnson D., Rohlfing T., Nelson J.,
RA   Stoneking T., Pepin K., Spieth J., Sekhon M., Armstrong J., Becker M.,
RA   Belter E., Cordum H., Cordes M., Courtney L., Courtney W., Dante M.,
RA   Du H., Edwards J., Fryman J., Haakensen B., Lamar E., Latreille P.,
RA   Leonard S., Meyer R., Mulvaney E., Ozersky P., Riley A., Strowmatt C.,
RA   Wagner-McPherson C., Wollam A., Yoakum M., Bell M., Dedhia N.,
RA   Parnell L., Shah R., Rodriguez M., Hoon See L., Vil D., Baker J.,
RA   Kirchoff K., Toth K., King L., Bahret A., Miller B., Marra M.,
RA   Martienssen R., McCombie W.R., Wilson R.K., Murphy G., Bancroft I.,
RA   Volckaert G., Wambutt R., Duesterhoeft A., Stiekema W., Pohl T.,
RA   Entian K.-D., Terryn N., Hartley N., Bent E., Johnson S.,
RA   Langham S.-A., McCullagh B., Robben J., Grymonprez B., Zimmermann W.,
RA   Ramsperger U., Wedler H., Balke K., Wedler E., Peters S.,
RA   van Staveren M., Dirkse W., Mooijman P., Klein Lankhorst R.,
RA   Weitzenegger T., Bothe G., Rose M., Hauf J., Berneiser S., Hempel S.,
RA   Feldpausch M., Lamberth S., Villarroel R., Gielen J., Ardiles W.,
RA   Bents O., Lemcke K., Kolesov G., Mayer K., Rudd S., Schoof H.,
RA   Schueller C., Zaccaria P., Mewes H.-W., Bevan M., Fransz P.;
RT   "Sequence and analysis of chromosome 5 of the plant Arabidopsis
RT   thaliana.";
RL   Nature 408:823-826(2000).
CC   -!- SIMILARITY: BELONGS TO THE NOP5/NOP56 FAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AF007271; AAB61073.1; -.
DR   InterPro; IPR002687; Nop.
DR   Pfam; PF01798; Nop; 1.
DR   ProDom; PD004104; Nop; 1.
KW   Hypothetical protein.
SQ   SEQUENCE   439 AA;  47933 MW;  BE0E3214F3E34FEB CRC64;

  Query Match           100.0%;  Score 6;  DB 1;  Length 439;
  Best Local Similarity   66.7%;  Pred. No. 47;
  Matches    4;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 WXXWHF 6
          |::|||
```

```
Db    184 WFGWHF 189


RESULT 6
FD6T_BRANA
ID   FD6C_BRANA     STANDARD;      PRT;   443 AA.
AC   P48627;
DT   01-FEB-1996 (Rel. 33, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Omega-6 fatty acid desaturase, chloroplast precursor (EC 1.14.99.-).
OS   Brassica napus (Rape).
OC   Eukaryota; Viridiplantae, Streptophyta, Embryophyta, Tracheophyta;
OC   Spermatophyta; Magnoliophyta, eudicotyledons, core eudicots, Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Brassica.
OX   NCBI_TaxID=3708;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Seed;
RX   MEDLINE=94340008; PubMed=8006133;
RA   Hitz W.D., Carlson T.J., Booth J.R. Jr., Kinney A.J., Stecca K.L.,
RA   Yadav N.S.;
RT   "Cloning of a higher-plant plastid omega-6 fatty acid desaturase cDNA
RT   and its expression in a cyanobacterium.";
RL   Plant Physiol. 105:635-641(1994).
CC   -!- FUNCTION: CHLOROPLAST OMEGA-6 FATTY ACID DESATURASE INTRODUCES
CC       THE SECOND DOUBLE BOND IN THE BIOSYNTHESIS OF 18:2 AND 18:3 FATTY
CC       ACIDS, IMPORTANT CONSTITUENTS OF PLANT MEMBRANES. IT IS THOUGHT
CC       TO USE FERREDOXIN AS AN ELECTRON DONOR AND IT ACTS ON FATTY ACIDS
CC       ESTERIFIED TO GALACTOLIPIDS, SULFOLIPIDS AND PHOSPHATIDYLGLYCEROL.
CC   -!- PATHWAY: Polyunsaturated fatty acid biosynthesis.
CC   -!- SUBCELLULAR LOCATION: CHLOROPLAST MEMBRANE PERIPHERAL.
CC   -!- DOMAIN: THE HISTIDINE BOX DOMAINS MAY CONTAIN THE ACTIVE SITE
CC       AND/OR BE INVOLVED IN METAL ION BINDING.
CC   -!- SIMILARITY: BELONGS TO THE FATTY ACID DESATURASE FAMILY.
CC   -!- SIMILARITY: CONTAINS 1 CYTOCHROME B5 HEME-BINDING DOMAIN.
CC   -!- SIMILARITY: CONTAINS 3 HISTIDINE BOXES.
CC   -!- SIMILARITY: CONTAINS 1 TRANSIT PEPTIDE.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L29214; AAA50107.1; -.
DR   InterPro; IPR001199; FA_desaturase.
DR   Pfam; PF00487; FA_desaturase; 1.
DR   ProDom; PD000181; FA_desaturase; 1.
KW   Oxidoreductase; Fatty acid biosynthesis; Chloroplast; Membrane;
KW   Transit peptide.
FT   TRANSIT       1     64       CHLOROPLAST (BY SIMILARITY).
FT   CHAIN        65    443       OMEGA-6 FATTY ACID DESATURASE.
FT   DOMAIN      160    170       HISTIDINE BOX 1.
FT   DOMAIN      202    206       HISTIDINE BOX-2.
FT   DOMAIN      362    366       HISTIDINE BOX-3.
SQ   SEQUENCE   443 AA,  50710 MW,  2B548B3AB17B11E CRC64;

  Query Match             Score 6;  Score/Ln:  6.0;  Length 44;
  Best Local Similarity   66.7%;  Pred. No. 48;
  Matches    4;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 WXXWHF 6
           |:::||
Db     252 WVNWHF 257


RESULT 7
FD6C_ARATH
ID   FD6C_ARATH     STANDARD;      PRT;   448 AA.
AC   P46312; Q9M094;
DT   01-NOV-1995 (Rel. 32, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Omega-6 fatty acid desaturase, chloroplast precursor (EC 1.14.99.-).
GN   FAD6 OR FADC OR AT4G30950 OR F6I18.140.
OS   Arabidopsis thaliana (Mouse ear cress).
OC   Eukaryota; Viridiplantae, Streptophyta, Embryophyta, Tracheophyta;
OC   Spermatophyta; Magnoliophyta, eudicotyledons, core eudicots, Rosidae;
OC   eurosids II; Brassicales, Brassicaceae, Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE OF 1-418 FROM N.A.
RC   STRAIN=cv. Columbia;
RX   MEDLINE=95148736; PubMed=7846158;
RA   Falcone D.L., Gibson S., Lemieux B., Somerville C.R.;
RT   "Isolation and complementation of Arabidopsis mutant
RT   deficient in chloroplast omega-6 desaturase activity.";
RL   Plant Physiol. 106:1453-1459(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Columbia;
RX   MEDLINE=20083488; PubMed=10617198;
RA   Mayer K.F.X., Schueller C., Wambutt R., Murphy G., Volckaert G.,
RA   Pohl T., Duesterhoeft A., Stiekema W., Entian K.-D., Terryn N.,
RA   Harris B., Ansorge W., Brandt P., Grivell L., Rieger M.,
RA   Weichselgartner M., de Simone V., Obermaier B., Mache R., Mueller M.,
RA   Kreis M., Delseny M., Puigdomenech P., Watson M., Schmidtheini T.,
RA   Reichert B., Portetelle D., Perez-Alonso M., Boutry M., Bancroft I.,
RA   Vos P., Hoheisel J., Zimmermann W., Wedler H., Ridley P.,
RA   Langham S.-A., McCullagh B., Bilham L., Robben J.,
RA   Van der Schueren J., Grymonprez B., Chuang Y.-J., Vandenbussche F.,
RA   Braeken M., Weltjens I., Voet M., Bastiaens I., Aert R., Defoor E.,
RA   Weitzenegger T., Bothe G., Ramsperger U., Hilbert H., Braun M.,
RA   Holzer E., Brandt A., Peters C., van Staveren M., Dirkse W.,
RA   Mooijman P., Klein Lankhorst R., Rose M., Hauf J., Koetter P.,
RA   Berneiser S., Hempel S., Feldpausch M., Lamberth S., Van den Daele H.,
RA   De Keyser A., Buysshaert C., Gielen J., Villarroel R., De Clercq R.,
RA   Van Montagu M., Rogers J., Cronin A., Quail M., Bray-Allen S.,
RA   Clark L., Doggett J., Hall S., Kay M., Lennard N., McLay K., Mayes R.,
RA   Pettett A., Rajandream M.A., Lyne M., Benes V., Rechmann S.,
RA   Borkova D., Bloecker H., Scharfe M., Grimm M., Loehnert T.-H.,
RA   Dose S., de Haan M., Maarse A., Schaefer M., Mueller-Auer S.,
RA   Gabel C., Fuchs M., Fartmann B., Granderath K., Dauner D., Herzl A.,
RA   Neumann S., Argiriou A., Vitale D., Liguori R., Piravandi E.,
RA   Massenet O., Quigley F., Clabauld G., Muendlein A., Felber R.,
RA   Schnabl S., Hiller R., Schmidt W., Lecharny A., Aubourg S.,
RA   Chefdor F., Cooke R., Berger C., Monfort A., Casacuberta E.,
RA   Gibbons T., Weber N., Vandenbol M., Bargues M., Terol J., Torres A.,
RA   Perez-Perez A., Purnelle B., Bent E., Johnson S., Tacon D., Jesse T.,
RA   Heijnen L., Schwarz S., Scholler P., Heber S., Francs P., Bielke C.,
RA   Frishman D., Haase D., Lemcke K., Mewes H.-W., Stocker S.,
RA   Zaccaria P., Bevan M., Wilson R.K., De la Bastide M., Habermann K.,
RA   Parnell L., Dedhia N., Gnoj L., Schutz K., Huang E., Spiegel L.,
RA   Sekhon M., Murray J., Sheet P., Cordes M., Abu-Threideh J.,
RA   Stoneking T., Kalicki J., Graves T., Harmon G., Edwards J.,
RA   Latreille P., Courtney L., Cloud J., Abbott A., Scott K., Johnson D.,
RA   Minx P., Bentley D., Fulton B., Miller N., Greco T., Kemp K.,
RA   Kramer J., Fulton L., Mardis E., Dante M., Pepin K., Hillier L.,
RA   Nelson J., Spieth J., Ryan E., Andrews S., Geletka C., Layman D.,
RA   Du H., Ali J., Berghoff A., Jones K., Drone K., Cotton M., Joshu C.,
RA   Antonoiu B., Zidanic M., Strong C., Sun H., Lamar B., Yordan C.,
RA   Ma P., Zhong J., Preston R., Vil D., Shekher M., Matero A., Shah R.,
RA   Swaby I.K., O'Shaughnessy A., Rodriguez M., Hoffman J., Till S.,
RA   Granat S., Shohdy N., Hasegawa A., Hameed A., Lodhi M., Johnson A.,
RA   Chen E., Marra M., Martienssen R., McCombie W.R.;
RT   "Sequence and analysis of chromosome 4 of the plant Arabidopsis
RT   thaliana.";
RL   Nature 402:769-777(1999).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Columbia;
RA   Shinozaki K., Davis R.W., Reiter J.P., Theologis A.;
RT   "RIKEN Arabidopsis full length cDNA clones (RAFLs) sequenced by the
RT   ABRC consortium (Salk/Stanford/PGEC).";
RL   Submitted (SEP-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: CHLOROPLAST OMEGA-6 FATTY ACID DESATURASE INTRODUCES
```

```
CC       THE SECOND DOUBLEBOND IN THE BIOSYNTHESIS OF 16.3 AND 18.3 FATTY
CC       ACIDS, IMPORTANT CONSTITUENTS OF PLANT MEMBRANES. IT IS THOUGHT
CC       TO USE FERREDOXIN AS AN ELECTRON DONOR AND TO ACT ON FATTY ACIDS
CC       ESTERIFIED TO GALACTOLIPIDS, SULFOLIPIDS AND PHOSPHATIDYLGLYCEROL.
CC   -!- PATHWAY: Polyunsaturated fatty acid biosynthesis.
CC   -!- SUBCELLULAR LOCATION: CHLOROPLAST, MEMBRANE-BOUND (PROBABLE).
CC   -!- DEVELOPMENTAL STAGE: HIGHEST LEVELS FOUND IN EXPANDING LEAVES.
CC   -!- DOMAIN: THE HISTIDINE BOX DOMAINS MAY CONTAIN THE ACTIVE SITE
CC       AND/OR BE INVOLVED IN METAL ION BINDING.
CC   -!- SIMILARITY: BELONGS TO THE FATTY ACID DESATURASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U09503; AAA92800.1; -.
DR   EMBL; AL022198; CAA18198.1; -.
DR   EMBL; AL161578; CAB79813.1; -.
DR   EMBL; AY045621; AAK73979.1; -.
DR   EMBL; AY058078; AAL24186.1; -.
DR   EMBL; AY058852; AAL24240.1; -.
DR   InterPro; IPR001225; FA_desaturase.
DR   Pfam; PF00487; FA_desaturase; 1.
DR   ProDom; PD001081; FA_desaturase; 1.
KW   Oxidoreductase; Fatty acid biosynthesis; Chloroplast; Membrane;
KW   Transit peptide.
FT   TRANSIT       1     59       CHLOROPLAST (BY SIMILARITY).
FT   CHAIN        70    448       OMEGA-6 FATTY ACID DESATURASE.
FT   DOMAIN      171    175       HISTIDINE BOX-1.
FT   DOMAIN      207    211       HISTIDINE BOX 2.
FT   DOMAIN      367    371       HISTIDINE BOX-3.
SQ   SEQUENCE   448 AA;  51229 MW;  C3AC72FB29FBF287 CRC64;

  Query Match           100.0%;  Score 6;  DB 1;  Length 448;
  Best Local Similarity   66.7%;  Pred. No. 48;
  Matches    4;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 WXXWHF 6
          |::|||
Db    257 WVNWHF 262


RESULT 8
CATA_PSEPU
ID   CATA_PSEPU     STANDARD;      PRT;   479 AA.
AC   Q59714;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   Catalase (EC 1.11.1.6).
GN   KATA OR CATA.
OS   Pseudomonas putida.
OC   Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC   Pseudomonas.
OX   NCBI_TaxID=303;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Corvallis;
RX   MEDLINE=98019091; PubMed=9358059;
RA   Kim Y.C., Miller C.D., Anderson A.J.;
RT   "Identification of adjacent genes encoding the major catalase and a
RT   bacterioferritin from the plant-beneficial bacterium Pseudomonas
RT   putida.";
RL   Gene 199:219-224(1997).
CC   -!- FUNCTION: DECOMPOSES HYDROGEN PEROXIDE IN WATER AND OXYGEN, SERVES
CC       TO PROTECT CELLS FROM THE TOXIC EFFECTS OF HYDROGEN PEROXIDE.
CC   -!- CATALYTIC ACTIVITY: 2 H(2)O(2) = O(2) + 2 H(2)O.
CC   -!- COFACTOR: HEME GROUP.
CC   -!- ENZYME REGULATION: ACTIVATED BY PEROXIDE.
CC   -!- SIMILARITY: BELONGS TO THE CATALASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U63511; AAB88219.1; -.
DR   HSSP; P42321; 2CAE.
DR   InterPro; IPR002226; Catalase.
DR   Pfam; PF00199; catalase; 1.
DR   PRINTS; PR00067; CATALASE.
DR   ProDom; PD000510; Catalase; 1.
DR   PROSITE; PS00437; CATALASE_1; 1.
DR   PROSITE; PS00438; CATALASE_2; 1.
KW   Oxidoreductase; Peroxidase; Iron; Heme; Hydrogen peroxide.
FT   ACT_SITE     53     53       BY SIMILARITY.
FT   ACT_SITE    126    126       BY SIMILARITY.
FT   BINDING     336    336       PROXIMAL HEME LIGAND (BY SIMILARITY).
SQ   SEQUENCE   479 AA;  53381 MW;  EFE3CBDE67778571 CRC64;

  Query Match           100.0%;  Score 6;  DB 1;  Length 479;
  Best Local Similarity   66.7%;  Pred. No. 50;
  Matches    4;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 WXXWHF 6
          |::|||
Db    209 WVKWHF 214


RESULT 9
SIK1_YEAST
ID   SIK1_YEAST     STANDARD;      PRT;   504 AA.
AC   Q12460;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   SIK1 protein (Nucleolar protein NOP56).
GN   SIK1 OR NOP56 OR YLR197W OR L8167.9.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN S288c / YPH1;
RX   MEDLINE=96040178; PubMed=7547500;
RA   Morin P.J., Downs J.A., Snodgrass A.M., Gilmore T.D.;
RT   "Genetic analysis of growth inhibition by GAI4 L kappa B alpha in
RT   Saccharomyces cerevisiae.";
RL   Cell Growth Differ. 6:789-798(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288c / AB972;
RA   Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Dover J.,
RA   Favello A., Fulton L., Gattung S., Greco T., Kirsten J., Kucaba T.,
RA   Hallsworth K., Hawkins J., Hillier L., Jier M., Johnson D.,
RA   Johnston L., Langston Y., Latreille P., Mardis E., Menezes S.,
RA   Miller N., Nhan M., Pauley A., Peluso D., Rifken L., Riles L.,
RA   Taich S., Trevaskis E., Vaudin M., Vignati D., Wilcox L., Wilson R.,
RA   Wohldman P., Waterston R.;
RL   Submitted (SEP-1994) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   CHARACTERIZATION, AND MUTAGENESIS.
RX   MEDLINE=98038777; PubMed=9372940;
RA   Gautier T., Berges T., Tollervey D., Hurt E.;
RT   "Nucleolar KKE/D repeat proteins Nop56p and Nop58p interact with Nop1p
RT   and are required for ribosome biogenesis.";
RL   Mol. Cell. Biol. 17:7088-7098(1997).
CC   -!- FUNCTION: REQUIRED FOR 60 S RIBOSOMAL SUBUNIT SYNTHESIS ...
```

```
CC   -!- SUBUNIT: INTERACTS WITH NOP1 AND NOP58.
CC   -!- SUBCELLULAR LOCATION: Nucleolar; nucleolar.
CC   -!- SIMILARITY: BELONGS TO THE NOP5/NOP58 FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a licensing agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U20037; AAC49066.1; -.
DR   EMBL; U14913; AAB67431.1; -.
DR   PIR; S61719; S61719.
DR   InterPro; IPR002687; Nop.
DR   Pfam; PF01798; Nop; 1.
DR   ProDom; PD004104; Nop; 1.
KW   Ribosome biogenesis; Nuclear protein.
FT   DOMAIN      443    504       ASP/GLU/LYS-RICH.
FT   MUTAGEN     333    333       V->A: REDUCED GROWTH RATE AT ALL
FT                                TEMPERATURES; WHEN ASSOCIATED WITH R-385.
FT   MUTAGEN     385    385       R->K: AT 37 DEGREES, GROWTH SLOWS AFTER 6
FT                                TO 8 HOURS AND CELL DIVISION STOPS AFTER
FT                                20 HOURS.
FT   MUTAGEN     385    385       R->K: REDUCED GROWTH RATE AT ALL
FT                                TEMPERATURES, WHEN ASSOCIATED WITH A-333.
SQ   SEQUENCE   614 AA,  69861 MW,  F9511A6973EF4942 CRC64;

  Query Match             [illegible]; Score 6; DB 1; Length 614;
  Best Local Similarity   66.7%; Pred. No. 53;
  Matches    4; Conservative    2; Mismatches    0; Indels    0; Gaps    0;

Qy       1 WXXWHF 6
           |::|'|
Db     129 WYGWHF 202


RESULT 10
NOP5_YEAST
ID   NOP5_YEAST     STANDARD;      PRT;   511 AA.
AC   Q12499;
DT   [illegible] (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15 JUN 2002 (Rel. 41, Last annotation update)
DE   Nucleolar protein NOP5 (Nucleolar protein NOP5).
GN   NOP5 OR NOP6 OR YOR310C OR O6119.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288c / FY1679;
RA   Feldmann H.M., Hegemann J., Wolf K., Philippsen P., Barlow M.,
RL   Submitted (AUG 1995) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A., AND CHARACTERIZATION.
RC   STRAIN=W303;
RX   MEDLINE=98356816; PubMed=9632810;
RA   Wu P., Brockenbrough J.S., Metcalfe A.C., Chen S., Aris J.P.,
RT   "Nop5p is a small nucleolar ribonucleoprotein component required for
RT   pre-18S rRNA processing in yeast.",
RL   J. Biol. Chem. 273:16453-16460(1998).
CC   -!- FUNCTION: REQUIRED FOR PRE-18S RRNA PROCESSING. MAY BIND
CC       MICROTUBULES.
CC   -!- SUBUNIT: INTERACTS WITH NOP1 AND NOP2.
CC   -!- SUBCELLULAR LOCATION: Nuclear; nucleolar.
CC   -!- SIMILARITY: BELONGS TO THE NOP5/NOP58 FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its

CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a licensing agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X99065; CAA67107.1; -.
DR   EMBL; Z75217; CAA99630.1; -.
DR   EMBL; AF086070; AAC39484.1; -.
DR   SGD; S0005837; NOP58.
DR   InterPro; IPR002687; Nop.
DR   Pfam; PF01798; Nop; 1.
DR   ProDom; PD004104; Nop; 1.
KW   Ribosome biogenesis; Nuclear protein; rRNA processing.
FT   DOMAIN      441    511       ASP/GLU/LYS-RICH.
SQ   SEQUENCE   511 AA,  [illegible] MW;  9A2989448E3A19E2 CRC64;

  Query Match             100.0%; Score 6; DB 1; Length 511;
  Best Local Similarity   66.7%; Pred. No. 53;
  Matches    4; Conservative    2; Mismatches    0; Indels    0; Gaps    0;

Qy       1 WXXWHF 6
           |::|'|
Db     197 WYGWHF 192


RESULT 11
NOP5_HUMAN
ID   NOP5_HUMAN     STANDARD;      PRT;   [illegible] AA.
AC   Q9Y2X3; Q9P036; Q9UFN3;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Nucleolar protein NOP5 (Nucleolar protein 5) (NOP58) (HSPC120).
GN   NOP5.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Lyman S.K., Gerace L.;
RT   "Cloning and characterization of NOP5/NOP58.";
RL   Submitted (JAN 1999) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20384617; PubMed=10995095;
RA   Nelson S.A., Cantura M.E., LaPorbella W.J.;
RT   "Identification and characterization of a novel PEDF-induced human gene.";
RL   Gene 255:87-96(2000).
RN   [3]
RP   SEQUENCE OF 1-447 FROM N.A.
RC   TISSUE=Brain;
RA   Bloecker H., Boecher M., Brandt P., Mewes H.-W., Gassenhuber J.,
RA   Wiemann S.;
RL   Submitted (SEP 1999) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE OF 64-529 FROM N.A.
RC   TISSUE=Blood;
RX   MEDLINE=20493367; PubMed=11042152;
RA   Zhang Q.-H., Ye M., Wu X.-Y., Ren S.-X., Zhao M., Zhao C.-J., Fu G.,
RA   Shen Y., Fan H.-Y., Lu G., Zhong M., Xu X.-R., Han Z.-G., Zhang J.-W.,
RA   Tao J., Huang Q.-H., Zhou J., Hu G.-X., Gu J., Chen S.-J., Chen Z.;
RT   "Cloning and functional analysis of cDNAs with open reading frames for
RT   [illegible] predicted genes expressed in fetal human hematopoietic
RT   stem/progenitor cells.",
RL   Genome Res. 10:1546-1560(2000).
CC   -!- FUNCTION: [illegible] (By
CC       similarity).
CC   -!- SUBCELLULAR LOCATION: Nuclear, nucleolar.
CC   -!- TISSUE SPECIFICITY: Ubiquitous.
CC   -!- SIMILARITY: BELONGS TO THE NOP5/NOP58 FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
```

```
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AF123534; AAD27610.1; -.
DR   EMBL; AF263608; AAF91394.1; -.
DR   EMBL; AL117554; CAB55989.1; -.
DP   EMBL; AF161469; AAF29084.1; -.
DR   SWISS-2DPAGE; Q9Y2X3; HUMAN.
DR   InterPro; IPR002687; Nop.
DR   Pfam; PF01798; Nop; 1.
DR   ProDom; PD004104; Nop; 1.
KW   Ribosome biogenesis; Nuclear protein.
FT   CONFLICT      2      2       L -> M (IN REF. 3).
FT   CONFLICT    129    129       G -> V (IN REF. 3).
FT   CONFLICT    202    221       LTYPFTLQKVGDPFNYASAF -> YHTASVYPFLAIGRLCL
FT                                CQ (IN REF. 4).
FT   CONFLICT    235    240       FAAAEISMGTEVSRRDICNILHLCTQ -> EGSCRDIHGNR
FT                                GPRRRYLQYSASLHP (IN REF. 4).
FT   CONFLICT    280    280       M -> V (IN REF. 3).
FT   CONFLICT    443    447       PKIFQ -> TQNTT (IN REF. 3).
SQ   SEQUENCE   529 AA;  59578 MW;  27CD73CFF5B9A556 CRC64;

  Query Match             100.0%;  Score 6;  DB 1;  Length 529;
  Best Local Similarity   66.7%;  Pred. No. 54;
  Matches    4,  Conservative    2,  Mismatches    0,  Indels    0,  Gaps    0;

Qy       1 WXXWHF 6
           |::|||
Db     186 WYGWHF 191


RESULT 12
NOP5_RAT
ID   NOP5_RAT       STANDARD;      PRT;   534 AA.
AC   Q9QZ86; O88525;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Nucleolar protein NOP5 (Nucleolar protein 5) (Nopp140-associated
DE   protein).
GN   NOP5 OR NAP65.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A., SEQUENCE OF 1-20, AND INTERACTION WITH NOLC1
RX   MEDLINE=20143579; PubMed=10679015;
RA   Yang Y., Isaac C., Wang C., Dragon F., Pogacic V., Meier U.T.;
RT   "Conserved composition of mammalian box H/ACA and box C/D small
RT   nucleolar ribonucleoprotein particles and their interaction with the
RT   common factor Nopp140.";
RL   Mol. Biol. Cell 11:567-577(2000).
RN   [2]
RP   SEQUENCE OF 1-461 FROM N.A.
RA   Hatton D., Gray J.C.;
RL   Submitted (JUN-1998) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: REQUIRED FOR 60S RIBOSOMAL SUBUNIT BIOGENESIS (By
CC       similarity).
CC   -!- SUBUNIT: Interacts with Nolc1/Nopp140.
CC   -!- SUBCELLULAR LOCATION: Nuclear, nucleolar.
CC   -!- SIMILARITY: BELONGS TO THE NOP5/NOP56 FAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AF194371; AAF05769.1; -.
DP   EMBL; AF069782; AAC23535.1; -.
DP   InterPro; IPR002687; Nop.
DR   Pfam; PF01798; Nop; 1.
DR   ProDom; PD004104; Nop; 1.
KW   Ribosome biogenesis; Nuclear protein.
FT   CONFLICT    396    396       R -> K (IN REF. 2).
FT   CONFLICT    459    459       A -> K (IN REF. 2).
SQ   SEQUENCE   534 AA;  60070 MW;  4B9585FA14E67799 CRC64;

  Query Match             100.0%;  Score 6;  DB 1;  Length 534;
  Best Local Similarity   66.7%;  Pred. No. 55;
  Matches    4,  Conservative    2,  Mismatches    0,  Indels    0,  Gaps    0;

Qy       1 WXXWHF 6
           |::|||
Db     186 WYGWHF 191


RESULT 13
Y4CD_RHISN
ID   Y4CD_RHISN     STANDARD;      PRT;   640 AA.
AC   P55386;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   Hypothetical 71.6 kDa protein Y4CD.
GN   Y4CD.
OS   Rhizobium sp. (strain NGR234).
OG   Plasmid sym pNGR234a.
OC   Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC   Rhizobiaceae; Rhizobium.
OX   NCBI_TaxID=394;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97305956; PubMed=9163424;
RA   Freiberg C.A., Fellay R., Bairoch A., Broughton W.J., Rosenthal A.,
RA   Perret X.;
RT   "Molecular basis of symbiosis between Rhizobium and legumes.";
RL   Nature 387:394-401(1997).
CC   -!- SIMILARITY: NONE EVIDOUS.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AE000067; AAB91634.1; -.
DR   InterPro; IPR000157; TIR_domain.
DR   SMART; SM00255; TIR; 1.
KW   Hypothetical protein; Plasmid.
SQ   SEQUENCE   640 AA;  71609 MW;  [illegible] CRC64;

  Query Match             100.0%;  Score 6;  DB 1;  Length 640;
  Best Local Similarity   66.7%;  Pred. No. 62;
  Matches    4,  Conservative    2,  Mismatches    0,  Indels    0,  Gaps    0;

Qy       1 WXXWHF 6
           |::|||
Db     521 WPFWHF 526


RESULT 14
PLSB_CAEEL
ID   PLSB_CAEEL     STANDARD;      PRT;   718 AA.
AC   Q22949;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
```

```
DT   15-JUL-1999 (Rel. 35, Last annotation update)
DE   Probable glycerol-3-phosphate acyltransferase, mitochondrial precursor
DE   (EC 2.3.1.15) (GPAT).
GN   F08F3.2.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditina;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Bristol N2;
RA   Blanchard M., Bradshaw H.;
RL   Submitted (JUL 1999) to the EMBL/GenBank/DDBJ databases.
CC   -!- CATALYTIC ACTIVITY: Acyl CoA + sn-glycerol 3 phosphate = CoA + 1
CC       acyl-sn-glycerol 3 phosphate.
CC   -!- PATHWAY: PHOSPHOLIPID BIOSYNTHESIS. TRANSFERASE; ACYLTRANSFERASE. IT MAY
CC       ALSO FUNCTION IN THE REGULATION OF MEMBRANE BIOGENESIS.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein. Mitochondrial
CC       (Potential).
CC   -!- SIMILARITY: BELONGS TO THE GPAT / DAPAT FAMILY.
CC   -----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; U64847; AAB04876.1; -.
DR   WormPep; F08F3.2; CE00267.
DR   InterPro; IPR002123; Acyltransferase.
DR   Pfam; PF01553; Acyltransferase; 1.
KW   Phospholipid biosynthesis; Transferase; Acyltransferase;
KW   Transmembrane; Mitochondrion; Transit peptide.
FT   TRANSIT       1      ?       MITOCHONDRION (POTENTIAL).
FT   CHAIN         ?    718       PROBABLE GLYCEROL 3 PHOSPHATE
FT                                ACYLTRANSFERASE.
FT   TRANSMEM    409    425       POTENTIAL.
SQ   SEQUENCE   718 AA;  82071 MW;  E0A36A4A86FC138D CRC64;

  Query Match             100.0%;  Score 6;  DB 1;  Length 718;
  Best Local Similarity    66.7%;  Pred. No. 63;
  Matches    4;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 WXXWHF 6
             |:: ||
Db     178 WCNWHF 183


RESULT 15
TXAI_ACTEQ
ID   TXAI_ACTEQ     STANDARD;      PRT;    82 AA.
AC   Q9NJQ2; Q9TXD2;
DT   15-JUN-2002 (Rel. 41, Created)
DT   15-JUN-2002 (Rel. 41, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Neurotoxin Ae I precursor (Toxin Ae I) (AeNa).
OS   Actinia equina (European sea anemone).
OC   Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia; Actiniaria;
OC   Nynantheae; Actiniidae; Actinia.
OX   NCBI_TaxID=6106;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20135764; PubMed=10669802;
RA   Anderluh G., Podlesek Z., Macek P.;
RT   "A cerebrotoxic motif in pigalutin of Cnidarian toxins and nematocyst
RT   collagens and its putative role.";
RL   Biochim. Biophys. Acta 1476:372-376(2000).
RN   [2]
RP   SEQUENCE OF 29-82.
RX   MEDLINE=96432268; PubMed=8835334;
RA   Lin X.-Y., Ishida M., Nagashima Y., Shiomi K.;
RT   "A polypeptide toxin in the sea anemone Actinia equina homologous with
RT   other sea anemone sodium channel toxins: isolation and amino acid
RT   sequence.";
RL   Toxicon 34:57-65(1996).
CC   -!- FUNCTION: Binds specifically to the sodium channel, thereby
CC       delaying its inactivation during signal transduction (By
CC       similarity). Causes death to crabs and mice.
CC   -!- SUBCELLULAR LOCATION: Cell wall and then secreted (By similarity).
CC   -!- SIMILARITY: BELONGS TO THE SEA ANEMONE SODIUM CHANNEL INHIBITORY
CC       TOXIN FAMILY.
CC   -----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; AF130344; AAF27538.1; -.
DR   HSSP; P01533; 1ATX.
DR   InterPro; IPR000633; Anemone_toxin.
DR   Pfam; PF00706; toxin_4; 1.
KW   Neurotoxin; Toxin; Sodium channel inhibitor; Signal; Cnidocyst.
FT   SIGNAL        1     19       POTENTIAL.
FT   PROPEP       20     28
FT   CHAIN        29     82       NEUROTOXIN AE I.
FT   DISULFID     32     79       BY SIMILARITY.
FT   DISULFID     34     69       BY SIMILARITY.
FT   DISULFID     62     80       BY SIMILARITY.
SQ   SEQUENCE   82 AA;  8695 MW;  D2B8A329AB86DF3D CRC64;

  Query Match             93.3%;  Score 5;  DB 1;  Length 82;
  Best Local Similarity   60.0%;  Pred. No. 1.7e+02;
  Matches    3;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 XXWHF 6
           ::|||
Db      64 KGWHF 68


Search completed: February 20, 2003, 11:04:57
Job time : 11.3333 secs
```

GenCore version 5.1.3

OM protein - protein search, using sw model

Run on: February 20, 2003, 11:23:27 ; Search Time 46 Seconds
(without alignments)
26.876 Million cell updates/sec

Title: US-09-912-414-11
Perfect score: 6
Sequence: 1 WXXWHF 6

Scoring table: OLIGODX
Gapop 60.0 , Gapext 60.0

Searched: 671580 seqs, 206047115 residues

Word size : 0

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : SPTREMBL_21:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	6	100.0	187	16	Q9CLS4	Q9cls4 pasteurella
2	6	100.0	195	4	Q96N05	Q96n05 homo sapien
3	6	100.0	201	10	Q9C600	Q9c600 arabidopsis
4	6	100.0	202	10	Q9AXY3	Q9axy3 brassica ca
5	6	100.0	213	10	Q9AXY5	Q9axy5 brassica na
6	6	100.0	216	10	Q9AXY2	Q9axy2 brassica na
7	6	100.0	219	10	Q94K47	Q94k47 arabidopsis
8	6	100.0	222	10	Q9AXY4	Q9axy4 brassica ol
9	6	100.0	240	16	O53604	O53604 mycobacteri
10	6	100.0	241	17	O59123	O59123 pyrococcus
11	6	100.0	256	10	Q9FIN2	Q9fin2 arabidopsis
12	6	100.0	260	16	O83250	O83250 treponema p
13	6	100.0	289	17	Q8TRL3	Q8trl3 methanosarc
14	6	100.0	293	16	Q91065	Q91065 pseudomonas
15	6	100.0	310	2	Q9EXF1	Q9exf1 listeria mo
16	6	100.0	314	16	Q9L120	Q9l120 streptomyce
17	6	100.0	381	2	Q93TQ4	Q93tq4 agrobacteri
18	6	100.0	388	16	Q8UB84	Q8ub84 agrobacteri
19	6	100.0	394	10	Q9AW80	Q9aw80 guillardia
20	6	100.0	403	16	Q8YF56	Q8yf56 brucella me
21	6	100.0	415	16	Q8RD64	Q8rd64 thermoanaer
22	6	100.0	421	10	Q9CBM4	Q9cbm4 chlamydomon
23	6	100.0	423	16	Q927H4	Q927h4 listeria in
24	6	100.0	423	16	Q8Y412	Q8y412 listeria mo
25	6	100.0	424	10	O48663	O48663 chlamydomon
26	6	100.0	452	16	Q8CL19	Q8cl19 salmonella
27	6	100.0	452	16	Q8Z2K1	Q8z2k1 salmonella
28	6	100.0	454	10	Q9ZRW0	Q9zrw0 cicer ariet
29	6	100.0	462	16	Q8X5T1	Q8x5t1 escherichia
30	6	100.0	472	16	Q8YMN8	Q8ymn8 anabaena sp
31	6	100.0	473	5	Q9BHU7	Q9bhu7 leishmania
32	6	100.0	473	11	O70396	O70396 mus musculu
33	6	100.0	475	17	Q9HSR8	Q9hsr8 halobacteri
34	6	100.0	485	10	O80401	O80401 oryza sativ
35	6	100.0	487	5	O45012	O45012 caenorhabdi
36	6	100.0	488	16	Q9A5V0	Q9a5v0 caulobacter
37	6	100.0	492	16	Q8YUZ1	Q8yuz1 anabaena sp
38	6	100.0	494	17	Q974C6	Q974c6 sulfolobus
39	6	100.0	497	3	O94514	O94514 schizosacch
40	6	100.0	499	10	Q9LTV0	Q9ltv0 arabidopsis
41	6	100.0	508	3	Q9P7S7	Q9p7s7 schizosacch
42	6	100.0	508	17	Q97WR4	Q97wr4 sulfolobus
43	6	100.0	510	5	Q9U5W4	Q9u5w4 drosophila
44	6	100.0	510	16	Q8XQC5	Q8xqc5 ralstonia s
45	6	100.0	511	5	Q9VM69	Q9vm69 drosophila

ALIGNMENTS

RESULT 1
Q9CLS4
ID Q9CLS4 PRELIMINARY; PRT; 187 AA.
AC Q9CLS4;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Hypothetical protein PM1135.
GN PM1135.
OS Pasteurella multocida.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Pasteurella.
OX NCBI_TaxID=747;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN PM70;
RX MEDLINE=21145866; PubMed=11248100;
RA May B.J., Zhang Q., Li L.L., Paustian M.L., Whittam T.S., Kapur V.;
RT "Complete genomic sequence of Pasteurella multocida Pm70.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).
CC -!- SIMILARITY: BELONGS TO THE TETR/ACRR FAMILY OF TRANSCRIPTIONAL
CC REGULATORS.
DR EMBL; AE006154; AAK03219.1; -.
DR InterPro; IPR001647; HTH_TetR.
DR Pfam; PF00440; tetR; 1.
DR PRINTS; PR00455; HTHTETR.
KW DNA-binding; Hypothetical protein; Transcription regulation;
KW Complete proteome.
SQ SEQUENCE 187 AA; 22325 MW; 28E3E3E5C1411115F CRC64;

Query Match 100.0%; Score 6; DB 16; Length 187;
Best Local Similarity 66.7%; Pred. No. 1.6e+02;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 WXXWHF 6
|::|||
Db 89 WNIWHF 94

```
RESULT 2
Q96N05
ID   Q96N05      PRELIMINARY;      PRT;   190 AA.
AC   Q96N05;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   cDNA FLJ10647 fis, clone NT2RI2009...
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Nishiyama K., Wagatsuma M., Ando K., Ando H., Yokoi T., Kodaira H.,
RA   Furuya T., Takahashi M., Kikkawa E., Omura Y., Abe K., Kamihara K.,
RA   Katsuta N., Sato K., Tanikawa M., Yamazaki M., Ogiyama T., Ito E.,
RA   Otsuki T., Sato H., Wakatsuki A., Ishii S., Yamamoto C., Isono F.,
RA   Kawai-Hio Y., Saito F., Nishikawa T., Kimura K., Yamashita H.,
RA   Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Murakawa K.,
RA   Kanehori K., Takahashi Fujii A., Oshima A., Ogiyama A., Kawabata B.,
RA   Suzuki Y., Sugano S., Nagahari K., Masuho Y., Nagai K., Isogai T.,
RT   "NEDO human cDNA sequencing project.";
RL   Submitted (OCT-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AK056171; BAB71110.1; -.
SQ   SEQUENCE   190 AA;  21014 MW;  B43B0FB1B14A27B7 CRC64;

  Query Match             100.0%;  Score 6;  DB 4;  Length 190;
  Best Local Similarity   66.7%;  Pred. No. 1.6e+02;
  Matches    4;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 WXXWHF 6
           |::|||
Db      82 WGKWHF 87


RESULT 3
Q9C600
ID   Q9C600      PRELIMINARY;      PRT;   201 AA.
AC   Q9C600;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   Omega 6 reductase (Fragment).
GN   FAD6.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. COLUMBIA;
RA   Fourmann M., Froger N., Brunel D.;
RT   "Amplified consensus gene markers: Tools designing for a genetic map
RT   of Arabidopsis-known-function genes in Brassica.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF229392; AAK00664.1; -.
DR   InterPro; IPR001225; FA_desaturase.
DR   Pfam; PF00487; FA_desaturase; 1.
DR   ProDom; PD000461; FA_desaturase; 1.
FT   NON_TER       1       1
FT   NON_TER     201     201
SQ   SEQUENCE   201 AA;  23...4 MW;  ...  CRC64;

  Query Match             100.0%;  Score 6;  DB 10;  Length 201;
  Best Local Similarity   66.7%;  Pred. No. 1.7e+02;
  Matches    4;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 WXXWHF 6
           |::|||
Db     156 WVNWHF 161


RESULT 4
Q9AXY3
ID   Q9AXY3      PRELIMINARY;      PRT;   202 AA.
AC   Q9AXY3;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Omega 6 reductase (Fragment).
GN   FAD6-BR-1.
OS   Brassica campestris (Field mustard).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Brassica.
OX   NCBI_TaxID=3711;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. R500;
RA   Fourmann M., Froger N., Brunel D.;
RT   "Amplified consensus gene markers: Tools designing for a genetic map
RT   of Arabidopsis known-function genes in Brassica.";
RL   Submitted (JAN-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF229390; AAK00662.1; -.
DR   InterPro; IPR001225; FA_desaturase.
DR   Pfam; PF00487; FA_desaturase; 1.
DR   ProDom; PD000461; FA_desaturase; 1.
FT   NON_TER       1       1
FT   NON_TER     202     202
SQ   SEQUENCE   202 AA;  ...  MW;  ...  CRC64;

  Query Match             100.0%;  Score 6;  DB 10;  Length 202;
  Best Local Similarity   66.7%;  Pred. No. 1.7e+02;
  Matches    4;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 WXXWHF 6
           |::|||
Db     157 WVNWHF 162


RESULT 5
Q9AXY5
ID   Q9AXY5      PRELIMINARY;      PRT;   213 AA.
AC   Q9AXY5;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Omega 6 reductase (Fragment).
GN   FAD6-BN-2.
OS   Brassica napus (Rape).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Brassica.
OX   NCBI_TaxID=3708;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. STELLAR;
RA   Fourmann M., Froger N., Brunel D.;
RT   "Amplified consensus gene markers: Tools designing for a genetic map
RT   of Arabidopsis known function genes in Brassica.";
RL   Submitted (JAN-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF229394; AAK00665.1; -.
DR   InterPro; IPR001225; FA_desaturase.
DR   Pfam; PF00487; FA_desaturase; 1.
DR   ProDom; PD000461; FA_desaturase; 2.
FT   NON_TER       1       1
FT   NON_TER     213     213
SQ   SEQUENCE   213 AA;  24...  MW;  ...EB4F95ADBDC06 CRC64;

  Query Match             100.0%;  Score 6;  DB 10;  Length 213;
  Best Local Similarity   66.7%;  Pred. No. 1.8e+02;
  Matches    4;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 WXXWHF 6
```

```
          |::|||
Db    154 WVNWHF 159


RESULT 6
Q9AXY2
ID   Q9AXY2      PRELIMINARY;      PRT;   216 AA.
AC   Q9AXY2;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Omega 6 reductase (Fragment).
GN   FAD6-BN-1.
OS   Brassica napus (Rape).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons, core eudicots, Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Brassica.
OX   NCBI_TaxID=3708;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. STELLAR;
RA   Fourmann M., Froger N., Brunel D.;
RT   "Amplified consensus gene markers. Tools designing for a genetic map
RT   of Arabidopsis-known-function genes in Brassica.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF229391; AAK00661.1; -.
DR   InterPro; IPR001225; FA_desaturase.
DR   Pfam; PF00487; FA_desaturase; 1.
DR   ProDom; PD001081; FA_desaturase; 2.
FT   NON_TER       1      1
FT   NON_TER     216    216
SQ   SEQUENCE   216 AA;  [illegible] MW;  [illegible] CRC64;

  Query Match             100.0%;  Score 6;  DB 10;  Length 216;
  Best Local Similarity   66.7%;  Pred. No. 1.8e+02;
  Matches    4;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 WXXWHF 6
          |::|||
Db     157 WVNWHF 162


RESULT 7
Q94K47
ID   Q94K47      PRELIMINARY;      PRT;   219 AA.
AC   Q94K47;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Putative 6-phosphogluconolactonase.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons, core eudicots, Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Yamada K., Lin S.X., Sakano H., Pham P.K., Banh J., Chung M.K.,
RA   Goldsmith A.D., Lee J.M., Quach H.L., Tang C.C., Teriumi M., Yu G.,
RA   Bowser L., Carninci P., Chen H., Cheuk R., Hayashizaki Y., Ishida J.,
RA   Jones T., Kamiya A., Karlin-Neumann G., Kawai J., Kim C., Koosema B.,
RA   Lam B., Lin J., Meyers M.C., Miranda M., Narusaka M., Nguyen M.,
RA   Palm C.J., Sakurai T., Saron M., Seki M., Shinn P., Southwick A.,
RA   Shinozaki K., Davis R.W., Ecker J.R., Theologis A.,
RT   "Full Length cDNA of gene K16H17.11/AT5g24400 (GI:10177901)";
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: CONTAINS 1 WD REPEAT (TRP-ASP DOMAIN).
DR   EMBL; AF370305; AAK44120.1; -.
DR   TIGRFAMs; TIGR01198; pgl; 1.
KW   Repeat; WD repeat.
SQ   SEQUENCE   219 AA;  24542 MW;  1994A723D2434B8A CRC64;

  Query Match             100.0%;  Score 6;  DB 10;  Length 219;
  Best Local Similarity   66.7%;  Pred. No. 1.8e+02;
  Matches    4;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 WXXWHF 6
          |::|||
Db    72 WARWHF 77


RESULT 8
Q9AXY4
ID   Q9AXY4      PRELIMINARY;      PRT;   222 AA.
AC   Q9AXY4;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Omega 6 reductase (Fragment).
GN   FAD6-BO-1.
OS   Brassica oleracea (Cauliflower).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons, core eudicots, Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Brassica.
OX   NCBI_TaxID=3712;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. RAPID CYCLING;
RA   Fourmann M., Froger N., Brunel D.;
RT   "Amplified consensus gene markers. Tools designing for a genetic map
RT   of Arabidopsis-known-function genes in Brassica.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF229389; AAK00661.1; -.
DR   InterPro; IPR001225; FA_desaturase.
DR   Pfam; PF00487; FA_desaturase; 1.
DR   ProDom; PD001081; FA_desaturase; 2.
FT   NON_TER       1      1
FT   NON_TER     222    222
SQ   SEQUENCE   222 AA;  25781 MW;  [illegible] CRC64;

  Query Match             100.0%;  Score 6;  DB 10;  Length 222;
  Best Local Similarity   66.7%;  Pred. No. 1.8e+02;
  Matches    4;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 WXXWHF 6
          |::|||
Db    163 WVNWHF 168


RESULT 9
O53604
ID   O53604      PRELIMINARY;      PRT;   230 AA.
AC   O53604;
DT   01-JUN-1998 (TrEMBLrel. 06, Created)
DT   01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Hypothetical protein Rv0059.
GN   RV0059 OR MTV030.02 OR MT0065.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Firmicutes; Actinobacteria, Actinobacteridae;
OC   Actinomycetales, Corynebacterineae, Mycobacteriaceae, Mycobacterium.
OX   NCBI_TaxID=1773;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37RV;
RX   MEDLINE=98295987; PubMed=9634230;
RA   Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA   Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaia F.,
RA   Badcock K., Basham D., Brown T., Chillingworth T., Connor R.,
RA   Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA   Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA   Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA   Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA   Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
PT   "Deciphering the biology of Mycobacterium tuberculosis from the
RT   complete genome sequence.";
```

```
RL   Nature [illegible] 544:1[illegible],.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=[illegible] / CSUH003H,
RA   Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA   Peterson J., DeBoy R., Dodson R., Durkin M., Haft D., Hickey E.,
RA   Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA   Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA   Bishai W.,
RT   "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT   laboratory strains.",
RL   Submitted (APR-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL021428; CAA16240.1; -.
DR   EMBL; AE000910; AAK44287.1; -.
DR   TIGR; MT0003; -.
DR   Tuberculist; Rv0389; -.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   230 AA;  25580 MW;  1A109D34B16690A8 CRC64;

  Query Match             100.0%;  Score 6;  DB 16;  Length 230;
  Best Local Similarity   66.7%;  Pred. No. 1.6e+02;
  Matches    4;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 WXXWHF 6
             |::|||
Db      26 WITWHF 31


RESULT 10
O59123
ID   O59123      PRELIMINARY;      PRT;   241 AA.
AC   O59123;
DT   01-AUG-1998 (TrEMBLrel. 07, Created)
DT   01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Hypothetical protein PH1454.
GN   PH1454.
OS   Pyrococcus horikoshii.
OC   Archaea; Euryarchaeota, Thermococci, Thermococcales, Thermococcaceae,
OC   Pyrococcus.
OX   NCBI_TaxID=53953;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=OT3;
RX   MEDLINE=98344137; PubMed=9679194;
RA   Kawarabayasi Y., Sawada M., Horikawa H., Haikawa Y., Hino Y.,
RA   Yamamoto S., Sekine M., Baba S., Kosugi H., Hosoyama A., Nagai Y.,
RA   Sakai M., Ogura K., Otsuka R., Nakazawa H., Takamiya M., Ohfuku Y.,
RA   Funahashi T., Tanaka T., Kudoh Y., Yamazaki J., Kushida N., Oguchi A.,
RA   Aoki K., Yoshizawa T., Nakamura Y., Robb F.T., Horikoshi K.,
RA   Masuchi Y., Shizuya H., Kikuchi H.,
RT   "Complete sequence and gene organization of the genome of a hyper
RT   therm[illegible]r Archaeon, Pyrococcus horikoshii OT3.",
RL   DNA Res. 5:55-76(1998).
DR   EMBL; AP000006; BAA30561.1; -.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   241 AA;  26494 MW;  A095C10EA1E70818 CRC64;

  Query Match             100.0%;  Score 6;  DB 17;  Length 241;
  Best Local Similarity   66.7%;  Pred. No. 1.6e+02;
  Matches    4;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 WXXWHF 6
             |::|||
Db     193 WQTWHF 198


RESULT 11
Q9FIN2
ID   Q9FIN2      PRELIMINARY;      PRT;   256 AA.
AC   Q9FIN2;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   21 JUN 2002 (TrEMBLrel. 21, Last annotation update)
DE   6-phosphoglucoslactonase-like protein.
OS   Arabidopsis thaliana (Mouse ear cress).
OC   Eukaryota, Viridiplantae, Streptophyta, Embryophyta, Tracheophyta;
OC   Spermatophyta, Magnoliophyta, eudicotyledons, core eudicots, Rosidae;
OC   [illegible]; Brassicales, Brassicaceae, Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=COLUMBIA;
RX   MEDLINE=99156233; PubMed=10049488;
RA   Artiguenave F., Wincker P., Brottier P., Nakamura Y., Katsuki H., Miyajima N.,
RA   Tabata S.;
RT   "Structural analysis of Arabidopsis thaliana chromosome 5. VIII.
RT   Sequence features of the regions of 1,081,958 bp covered by seventeen
RT   physically assigned P1 and TAC clones.";
RL   DNA Res. 5:379-391(1998).
DR   EMBL; AB016884; BAB11233.1; -.
DR   InterPro; IPR000457; Glucosamine_iso.
DR   Pfam; PF01182; Glucosamine_iso; 1.
DR   TIGRFAMs; TIGR01198; pgl; 1.
DR   PROTEIN; 256 AA;  28734 MW;  21F7946B7BD770AF CRC64;

  Query Match             100.0%;  Score 6;  DB 10;  Length 256;
  Best Local Similarity   66.7%;  Pred. No. 2e+02;
  Matches    4;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 WXXWHF 6
             |::|||
Db      72 WARWHF 77


RESULT 12
O83250
ID   O83250      PRELIMINARY;      PRT;   260 AA.
AC   O83250;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   21 JUN 2002 (TrEMBLrel. 21, Last annotation update)
DE   D,D-carboxypeptidase, putative.
GN   TP0221.
OS   Treponema pallidum.
OC   Bacteria, Spirochaetales, Spirochaetaceae; Treponema.
OX   NCBI_TaxID=160;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NICHOLS;
RX   MEDLINE=98055876; PubMed=9405370;
RA   Fraser C.M., Norris S.J., Weinstock G.M., White O., Sutton G.G.,
RA   Dodson R., Gwinn M., Hickey E.K., Clayton R., Ketchum K.A.,
RA   Sodergren E., Hardham J.M., McLeod M.P., Salzberg S., Peterson J.,
RA   Khalak H., Richardson D., Howell J.K., Chidambaram M., Utterback T.,
RA   McDonald L., Artiach P., Bowman C., Cotton M.D., Fujii C., Garland S.,
RA   Hatch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O.,
RA   Venter J.C.;
RT   "Complete genome sequence of Treponema pallidum, the syphilis
RT   spirochete.";
RL   Science 281:375-388(1998).
DR   EMBL; AF001208; AAC65208.1; -.
DR   TIGR; TP0221; -.
DR   InterPro; IPR000561; EGF-like.
DR   InterPro; IPR003709; VanY.
DR   Pfam; PF02557; VanY; 1.
DR   PROSITE; PS00022; EGF_1; UNKNOWN_1.
KW   Carboxypeptidase; Complete proteome.
SQ   SEQUENCE   260 AA;  29193 MW;  DB5801245133B556 CRC64;

  Query Match             100.0%;  Score 6;  DB 16;  Length 260;
  Best Local Similarity   66.7%;  Pred. No. 2.1e+02;
  Matches    4;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 WXXWHF 6
             |::|||
```

```
Db    220 WEPWHF 225


RESULT 13
Q8TRL3
ID   Q8TRL3      PRELIMINARY;      PRT;   294 AA.
AC   Q8TRL3;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Hypothetical protein MA1162.
GN   MA1162.
OS   Methanosarcina acetivorans.
OC   Archaea; Euryarchaeota; Methanococci; Methanosarcinales;
OC   Methanosarcinaceae; Methanosarcina.
OX   NCBI_TaxID=2214;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C2A / ATCC 35395 / DSM 2834;
RX   MEDLINE=21929760; PubMed=11932238;
RA   Galagan J.E., Nusbaum C., Roy A., Endrizzi M.G., MacDonald P.,
RA   FitzHugh W., Calvo S., Engels R., Smirnov S., Atnoor D., Brown A.,
RA   Allen N., Naylor J., Stange-Thomann N., DeArellano K., Johnson R.,
RA   Linton L., McEwan P., McKernan K., Talamas J., Tirrell A., Ye W.,
RA   Zimmer A., Barber R.D., Cann I., Graham D.E., Grahame D.A., Guss A.M.,
RA   Hedderich R., Ingram-Smith C., Kuettner H.C., Krzycki J.A.,
RA   Leigh J.A., Li W., Liu J., Mukhopadhyay B., Reeve J.N., Smith K.,
RA   Springer T.A., Umayam L.A., White O., White R.H., de Macario E.C.,
RA   Ferry J.G., Jarrell K.F., Jing H., Macario A.J.L., Paulsen I.,
RA   Pritchett M., Sowers K.R., Swanson R.V., Zinder S.H., Lander E.,
RA   Metcalf W.W., Birren B.;
RT   "The genome of Methanosarcina acetivorans reveals extensive metabolic
RT   and physiological diversity.";
RL   Genome Res. 12:532-542(2002).
DR   EMBL; AE010783; AAM04683.1; -.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   289 AA;  32289 MW;  B4221708EA52BE66 CRC64;

  Query Match             100.0%;  Score 6;  DB 17;  Length 289;
  Best Local Similarity   66.7%;  Pred. No. 2.2e+02;
  Matches    4;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 WXXWHF 6
           |::|||
Db     168 WSLWHF 173


RESULT 14
Q9I065
ID   Q9I065      PRELIMINARY;      PRT;   293 AA.
AC   Q9I065;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Hypothetical protein PA2778.
GN   PA2778.
OS   Pseudomonas aeruginosa.
OC   Bacteria, Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC   Pseudomonas.
OX   NCBI_TaxID=287;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 15692 / PAO1;
RX   MEDLINE=20437337; PubMed=10984043;
RA   Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warrener P.,
RA   Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA   Garber R.L., Goltry L., Tolentino E., Westbrock-Wadman S., Yuan Y.,
RA   Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA   Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA   Reizer J., Saier M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT   "Complete genome sequence of Pseudomonas aeruginosa PAO1, an
RT   opportunistic pathogen.";
RL   Nature 406:959-964(2000).
DR   EMBL; AE004795; AAG06166.1; -.
DR   InterPro; IPR005074; Peptidase_C39.
DR   InterPro; IPR001440; TPR.
DR   Pfam; PF03412; Peptidase_C39; 1.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   293 AA,  32016 MW;  06A4101D6ED760E1 CRC64;

  Query Match             100.0%;  Score 6;  DB 15;  Length 293;
  Best Local Similarity   66.7%;  Pred. No. 2.3e+02;
  Matches    4;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 WXXWHF 6
           |::|||
Db     101 WPRWHF 106


RESULT 15
Q9EXF1
ID   Q9EXF1      PRELIMINARY;      PRT;   310 AA.
AC   Q9EXF1;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Putative phosphotransferase D-arabitol specific component IIC
DE   (Fragment).
GN   ATLC.
OS   Listeria monocytogenes.
OC   Bacteria, Firmicutes, Bacillus/Clostridium group, Bacillales;
OC   Listeriaceae; Listeria.
OX   NCBI_TaxID=1639;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=LO28;
RX   MEDLINE=21212979, PubMed=11316371;
RA   Saklani-Jusforgues H., Fontan E., Goossens P.L.;
RT   "Characterisation of a Listeria monocytogenes mutant deficient in D-
RT   arabitol fermentation.",
RL   Res. Microbiol. 152:175-177(2001).
DR   EMBL; AJ292552; CAC20641.1; -.
DR   InterPro; IPR004703, Gal_spec_IIC.
DR   Pfam; PF03611; EIIC-GAT; 1.
KW   Transferase.
FT   NON_TER       1      1
FT   NON_TER     310    310
SQ   SEQUENCE   310 AA;  33711 MW;  A2F9FE9CBC88DACF CRC64;

  Query Match             100.0%;  Score 6;  DB 2;  Length 310;
  Best Local Similarity   66.7%;  Pred. No. 2.4e+02;
  Matches    4;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 WXXWHF 6
           |::|||
Db      18 WNYWHF 23


Search completed: February 20, 2003, 11:28:57
Job time : 46 secs
```

GenCore version 5.1.3

OM protein - protein search, using sw model

Run on: February 20, 2003, 11:21:07 ; Search time 29.3333 Seconds
(without alignments)
27.256 Million cell updates/sec

Title: US-09-912-414-11
Perfect score: 6
Sequence: 1 WXXWHF 6

Scoring table: OLIGODX
Gapop 60.0 , Gapext 60.0

Searched: 908470 seqs, 133250620 residues

Word size : 0

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : A_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	6	100.0	6	21	AAB01492	Peptide which bind
2	6	100.0	6	21	AAB01499	Peptide which bind
3	6	100.0	6	21	AAB01505	Peptide which bind
4	6	100.0	6	21	AAB01506	Peptide which bind
5	6	100.0	36	22	ABB40893	Peptide #8399 enco
6	6	100.0	36	22	AAM61753	Human brain expres
7	6	100.0	36	22	AAM74549	Human bone marrow
8	6	100.0	36	22	AAM44663	Peptide #8700 enco
9	6	100.0	43	22	AAU14683	Novel bone marrow
10	6	100.0	53	23	ABG64892	Human albumin fusi
11	6	100.0	53	23	AAE21233	Human gene 18 enco
12	6	100.0	57	22	AAU56179	Propionibacterium
13	6	100.0	66	23	ABB07622	Human CREX protein
14	6	100.0	87	22	AAU56391	Propionibacterium
15	6	100.0	96	22	AAU14777	Novel bone marrow
16	6	100.0	110	23	ABB06130	Human CREX protein
17	6	100.0	173	22	AAG80884	Lipid modification
18	6	100.0	173	22	AAG83334	P patens lipid met
19	6	100.0	175	22	ABB03211	Human musculoskele
20	6	100.0	175	22	AAM42371	Human polypeptide
21	6	100.0	256	21	AAG39463	Arabidopsis thalia
22	6	100.0	256	21	AAG44075	Arabidopsis thalia
23	6	100.0	256	21	AAG48203	Arabidopsis thalia
24	6	100.0	259	22	AAG33113	C glutamicum prote
25	6	100.0	306	21	AAG05810	Arabidopsis thalia
26	6	100.0	306	21	AAG47194	Arabidopsis thalia
27	6	100.0	324	21	AAG50316	Arabidopsis thalia
28	6	100.0	325	21	AAG39462	Arabidopsis thalia
29	6	100.0	325	21	AAG44074	Arabidopsis thalia
30	6	100.0	325	21	AAG48202	Arabidopsis thalia
31	6	100.0	338	21	AAG05809	Arabidopsis thalia
32	6	100.0	338	21	AAG47193	Arabidopsis thalia
33	6	100.0	376	19	AAW46287	Rhizobium meliloti
34	6	100.0	377	19	AAW46288	Brucella abortus D
35	6	100.0	418	19	AAR69501	Linoleic-acid-desa
36	6	100.0	423	21	AAG50315	Arabidopsis thalia
37	6	100.0	423	23	ABB49593	Listeria monocyto
38	6	100.0	424	22	AAE13431	Soybean microsomal
39	6	100.0	427	22	AAU20451	Human secreted pro
40	6	100.0	427	22	AAU21833	Novel human neopla
41	6	100.0	448	21	AAG05808	Arabidopsis thalia
42	6	100.0	448	21	AAG47192	Arabidopsis thalia
43	6	100.0	448	23	ABB03179	Herbicidally activ
44	6	100.0	462	21	AAG44073	Arabidopsis thalia
45	6	100.0	466	22	AAB05611	Human protein sequ

ALIGNMENTS

RESULT 1
AAB01492
ID AAB01492 standard; peptide; 6 AA.
XX
AC AAB01492;
XX
DT 08-NOV-2000 (first entry)
XX
DE Peptide which binds to transcription factor E2F 1 DNA binding domain.
XX
KW DNA binding; transcription factor, E2F, E2F 1, cell cycle; DP 1;
KW activation; transcription; apoptosis; proliferative disorder;
KW psoriasis; restenosis.
XX
OS Synthetic.
XX
PN WO200044771-A1.
XX
PD 03-AUG-2000.
XX
PF 26-JAN-2000; 2000WO-GB00227
XX
PR 26-JAN-1999; 99GB-0001710.
XX
PA (PROL-) PROLIFIX LTD.
XX
PI Mueller R, Kontermann RE, Montigiani S;
XX
DR WPI; 2000-532806/48.
XX
PT Peptides binding to the DNA binding domain of transcription factor E2F
PT and inhibiting cell cycle progression, useful for the treatment of
PT cancer

```
XX
PS   Claim 6; Page 2; 42pp; English.
XX
CC   Peptides which bind to the DNA binding domain of transcription
CC   factor E2F and inhibit cell cycle progression may be useful as
CC   research agents to investigate the interaction between E2F and DP-1,
CC   or the activation of transcription by E2F 1/DP 1 heterodimers. They
CC   may also be used for inducing apoptosis and/or cell cycle arrest in
CC   a cell, particularly for treatment of cancer or other proliferative
CC   disorders such as psoriasis and restenosis.
XX
SQ   Sequence   6 AA;

  Query Match             100.0%;  Score 6;  DB 21;  Length 6;
  Best Local Similarity   66.7%;  Pred. No. 7.8e+05;
  Matches    4;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 WXXWHF 6
           |::|||
Db       1 WVRWHF 6


RESULT 2
AAB01499
ID   AAB01499 standard; peptide; 6 AA
XX
AC   AAB01499;
XX
DT   08-NOV-2000  (first entry)
XX
DE   Peptide which binds to transcription factor E2F 1 DNA binding domain
XX
KW   DNA binding; transcription factor; E2F; E2F 1; cell cycle; DP 1;
KW   activation; transcription; apoptosis; proliferative disorder;
KW   psoriasis; restenosis.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 2
FT                   /note= "Any amino acid"
FT   Misc-difference 3
FT                   /note= "Any amino acid"
XX
PN   WO200044771-A1.
XX
PD   03-AUG-2000
XX
PF   26 JAN 2000, [illegible]
XX
PR   26 JAN 1999,   [illegible]
XX
PA   (PROL-) PROLIFIX LTD.
XX
PI   Mueller R,  Kontermann RE,  Montigiani S;
XX
DR   WPI; 2000-530806/48
XX
PT   Peptide [illegible] DNA binding domain of transcription factor E2F
PT   and inhibiting cell cycle progression, useful for the treatment of
PT   cancer
XX
PS   Claim 4; Page 9; 42pp; English.
XX
CC   Peptides which bind to the DNA binding domain of transcription
CC   factor E2F and inhibit cell cycle progression may be useful as
CC   research agents to investigate the interaction between E2F and DP-1,
CC   or the activation of transcription by E2F 1/DP 1 heterodimers. They
CC   may also be used for inducing apoptosis and/or cell cycle arrest in
CC   a cell, particularly for treatment of cancer or other proliferative
CC   disorders such as psoriasis and restenosis.
XX
SQ   Sequence   6 AA;

  Query Match             100.0%;  Score 6;  DB 21;  Length 6;
  Best Local Similarity   66.7%;  Pred. No. 7.8e+05;
  Matches    4;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 WXXWHF 6
           |::|||
Db       1 WVRWHF 6


RESULT 3
AAB01505
ID   AAB01505 standard; peptide; 6 AA.
XX
AC   AAB01505;
XX
DT   08-NOV-2000  (first entry)
XX
DE   Peptide which binds to transcription factor E2F-1 DNA binding domain.
XX
KW   DNA binding; transcription factor; E2F; E2F-1; cell cycle; DP-1;
KW   activation; transcription; apoptosis; proliferative disorder;
KW   psoriasis; restenosis.
XX
OS   Synthetic.
XX
PN   WO200044771-A1.
XX
PD   03-AUG-2000.
XX
PF   26 JAN 2000, 2000WO-GB00227.
XX
PR   26 JAN 1999,   99GB-0001710.
XX
PA   (PROL-) PROLIFIX LTD.
XX
PI   Mueller R,  Kontermann RE,  Montigiani S;
XX
DR   WPI; 2000-530806/48.
XX
PT   Peptides binding to the DNA binding domain of transcription factor E2F
PT   and inhibiting cell cycle progression, useful for the treatment of
PT   cancer
XX
PS   Example; Page 26; 42pp; English.
XX
CC   Peptides which bind to the DNA binding domain of transcription
CC   factor E2F and inhibit cell cycle progression may be useful as
CC   research agents to investigate the interaction between E2F and DP-1,
CC   or the activation of transcription by E2F 1/DP 1 heterodimers. They
CC   may also be used for inducing apoptosis and/or cell cycle arrest in
CC   a cell, particularly for treatment of cancer or other proliferative
CC   disorders such as psoriasis and restenosis.
XX
SQ   Sequence   6 AA;

  Query Match             100.0%;  Score 6;  DB 21;  Length 6;
  Best Local Similarity   66.7%;  Pred. No. 7.8e+05;
  Matches    4;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 WXXWHF 6
           |::|||
Db       1 WARWHF 6


RESULT 4
AAB01506
ID   AAB01506 standard; peptide; 6 AA.
XX
AC   AAB01506;
XX
DT   08 NOV 2000  (first entry)
XX
```

```
DE   Peptide which binds to transcription factor E2F-1 DNA binding domain
XX
KW   DNA binding; transcription factor, E2F, E2F-1, cell cycle, DP-1;
KW   activation; transcription; apoptosis; proliferative disorder;
KW   psoriasis; restenosis.
XX
OS   Synthetic.
XX
PN   WO200044771-A1.
XX
PD   03-AUG-2000.
XX
PF   26-JAN-2000; 2000WO-GB00227.
XX
PR   26-JAN-1999;   99GB-0001710.
XX
PA   (PROL-) PROLIFIX LTD.
XX
PI   Mueller R,  Kontermann RE,  Montigiani S;
XX
DR   WPI; 2000-532806/48
XX
PT   Peptides binding to the DNA binding domain of transcription factor E2F
PT   and inhibiting cell cycle progression, useful for the treatment of
PT   cancer
XX
PS   Example; Page 26; 42pp; English.
XX
CC   Peptides which bind to the DNA binding domain of transcription
CC   factor E2F and inhibit cell cycle progression may be useful as
CC   research agents to investigate the interaction between E2F and DP-1,
CC   or the activation of transcription by E2F-1/DP-1 heterodimers. They
CC   may also be used for inducing apoptosis and/or cell cycle arrest in
CC   a cell, particularly for treatment of cancer or other proliferative
CC   disorders such as psoriasis and restenosis.
XX
SQ   Sequence   6 AA;

  Query Match             100.0%;  Score 6;  DB 21;  Length 6;
  Best Local Similarity   66.7%;  Pred. No. 7.8e+05;
  Matches    4;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 WXXWHF 6
          |::|||
Db      1 WVAWHF 6


RESULT 5
ABB40893
ID   ABB40893 standard; Peptide; 36 AA.
XX
AC   ABB40893;
XX
DT   04-FEB-2002  (first entry)
XX
DE   Peptide #8399 encoded by human foetal liver single exon probe.
XX
KW   Human; foetal liver; gene expression, single exon nucleic acid probe.
XX
OS   Homo sapiens
XX
PN   WO200157277-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US00669.
XX
PR   04-FEB-2000; 2000US-0180312.
PR   26-MAY-2000; 2000US-0207456.
PR   30-JUN-2000; 2000US-0608408.
PR   03-AUG-2000; 2000US-0632366.
PR   21-SEP-2000; 2000US-0234687.
PR   27-SEP-2000; 2000US-0236359.
PR   04-OCT-2000; 2000GB-0024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR   WPI; 2001-483447/52.
XX
PT   Human genome derived single exon nucleic acid probes useful for
PT   analyzing gene expression in human fetal liver -
XX
PS   Claim 27, SEQ ID NO 33528, 639pp + sequence listing, English.
XX
CC   The invention relates to a single exon nucleic acid probe for
CC   measuring human gene expression in a sample derived from human foetal
CC   liver. The single exon nucleic acid probes may be used for predicting,
CC   measuring and displaying gene expression in samples derived from human
CC   fetal liver  The present sequence is a peptide encoded by a single exon
CC   nucleic acid probe of the invention.
CC   Note: The sequence data for this patent did not form part of the
CC   printed specification, but was obtained in electronic format directly
CC   from WIPO at ftp wipo int/pub/published_pct_sequences.
XX
SQ   Sequence   36 AA;

  Query Match             100.0%;  Score 6;  DB 22;  Length 36;
  Best Local Similarity   66.7%;  Pred. No. 25;
  Matches    4;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 WXXWHF 6
          |::|||
Db     19 WRAWHF 24


RESULT 6
AAM61753
ID   AAM61753 standard; Protein; 36 AA.
XX
AC   AAM61753;
XX
DT   05-NOV-2001  (first entry)
XX
DE   Human brain expressed single exon probe encoded protein SEQ ID NO: 33858.
XX
KW   Human; brain expressed exon; gene expression analysis; probe;
KW   microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
KW   epilepsy; cancer.
XX
OS   Homo sapiens.
XX
PN   WO200157275-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US00667.
XX
PR   04-FEB-2000; 2000US-0180312.
PR   26-MAY-2000; 2000US-0207456.
PR   30-JUN-2000; 2000US-0608408.
PR   03-AUG-2000; 2000US-0632366.
PR   21-SEP-2000; 2000US-0234687.
PR   27-SEP-2000; 2000US-0236359.
PR   04-OCT-2000; 2000GB-0024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR   WPI; 2001-483446/52.
XX
PT   Single exon nucleic acid probes for analyzing gene expression in human
PT   brains -
XX
```

```
PS   Example 4; SEQ ID NO: 33858; 650pp + Sequence Listing; English.
XX
CC   The present invention provides a number of single exon nucleic acid
CC   probes which are derived from genomic sequences expressed in the human
CC   brain. They can be used to measure gene expression in brain cell samples,
CC   which may enable the diagnosis and improved treatment of nervous system
CC   diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
CC   epilepsy and cancers. The present sequence is a protein encoded by one of
CC   the probes of the invention.
XX
SQ   Sequence   36 AA;

  Query Match             100.0%;  Score 6;  DB 22;  Length 36;
  Best Local Similarity   66.7%;  Pred. No. 25;
  Matches    4;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 WXXWHF 6
           |::|||
Db      19 WRAWHF 24


RESULT 7
AAM74549
ID   AAM74549 standard; Protein; 36 AA.
XX
AC   AAM74549;
XX
DT   06-NOV-2001  (first entry)
XX
DE   Human bone marrow expressed probe encoded protein SEQ ID NO: 34855.
XX
KW   Human, bone marrow expressed exon, gene expression analysis, probe,
KW   microarray, cancer, leukaemia, lymphoma, myeloma.
XX
OS   Homo sapiens.
XX
PN   WO200157276-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US00668.
XX
PR   04-FEB-2000; 2000US-0180312.
PR   26-MAY-2000; 2000US-0207456.
PR   30-JUN 2000; 2000US-0609408.
PR   03-AUG-2000; 2000US-0632366.
PR   21-SEP-2000; 2000US-0234687.
PR   27-SEP-2000; 2000US-0676164.
PR   04-OCT-2000; 2000GB-0024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR   WPI; 2001-488900/53.
XX
PT   Human genome-derived single exon nucleic acid probes useful for
PT   analysing gene expression in human b cell matrix
XX
PS   Example 4; SEQ ID NO: 34855; 650pp + Sequence Listing; English.
XX
CC   The present invention provides a number of single exon nucleic acid
CC   probes which are derived from genomic sequences expressed in the human
CC   bone marrow. They can be used to measure gene expression in bone marrow
CC   samples, which may enable the improved diagnosis and treatment of cancers
CC   such as lymphoma, leukaemia and myeloma. The present sequence is a
CC   protein encoded by one of the probes of the invention.
XX
SQ   Sequence   36 AA;

  Query Match             100.0%;  Score 6;  DB 22;  Length 36;
  Best Local Similarity   66.7%;  Pred. No. 25;
  Matches    4;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 WXXWHF 6
           |::|||
Db      19 WRAWHF 24


RESULT 8
AAM34663
ID   AAM34663 standard; Protein; 36 AA.
XX
AC   AAM34663;
XX
DT   17-OCT-2001  (first entry)
XX
DE   Peptide #8700 encoded by probe for measuring placental gene expression.
XX
KW   Probe, microarray, human, placenta, antenatal diagnosis;
KW   genetic disorder.
XX
OS   Homo sapiens.
XX
PN   WO200157272-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US00663.
XX
PR   04-FEB-2000; 2000US-0180312.
PR   26-MAY-2000; 2000US-0207456.
PR   30-JUN 2000; 2000US-0609408.
PR   03-AUG-2000; 2000US-0632366.
PR   21-SEP-2000; 2000US-0234687.
PR   27-SEP-2000; 2000US-0236359.
PR   04-OCT-2000; 2000GB-0024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR   WPI; 2001-488897/53.
XX
PT   Human genome derived single exon nucleic acid probes useful for
PT   analyzing gene expression in human placenta -
XX
PS   Claim 27; SEQ ID No 34932; 654pp; English.
XX
CC   The present invention relates to single exon nucleic acid probes (SENP)
CC   see AA11715-AA15754. The present sequence is a peptide encoded by one
CC   such probe. The probes are useful for producing a microarray for
CC   predicting, measuring and displaying gene expression in samples derived
CC   from human placenta. The probes are useful for antenatal diagnosis of
CC   human genetic disorders.
XX
SQ   Sequence   36 AA;

  Query Match             100.0%;  Score 6;  DB 22;  Length 36;
  Best Local Similarity   66.7%;  Pred. No. 25;
  Matches    4;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 WXXWHF 6
           |::|||
Db      19 WRAWHF 24


RESULT 9
AAU14683
ID   AAU14683 standard; Protein; 43 AA
XX
AC   AAU14683;
XX
DT   24-OCT-2001  (first entry)
XX
DE   Novel bone marrow polypeptide #82.
```

```
XX
KW   Bone marrow; diagnostic; therapeutic; gene therapy; antigenic;
KW   haematopoiesis; myeloid; lymph cell disorder; tissue regeneration;
KW   wound healing; nutritional supplement; immune disorder;
KW   severe combined immunodeficiency; SCID.
XX
OS   Homo sapiens.
XX
PN   WO200157187-A2
XX
PD   09-AUG-2001.
XX
PF   05-FEB-2001; 2001WO-US03782
XX
PR   03-FEB-2000; 2000US-0496914.
PR   20-JUN-2000; 2000US-0598075.
PR   19-JUL-2000; 2000US-0620325.
PR   30-NOV-2000; 2000US-0250683.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Ford JE,  Boyle BJ,  Tang YT,  Liu C,  Asundi V,  Zhou P,  Xue AJ,
PI   Ren F,  Drmanac RT;
XX
DR   WPI; 2001-488875/53.
DR   N-PSDB; AAS22988
XX
PT   Nucleic acids encoding bone marrow polypeptides, useful in diagnostic
PT   and gene therapy -
XX
PS   Claim 10; Page 248; 392pp; English.
XX
CC   AAU14602-AAU14794 represent novel bone marrow polypeptides of the
CC   invention. The proteins and corresponding coding sequences may be used
CC   in the prevention, diagnosis and treatment of diseases associated with
CC   inappropriate bone marrow polypeptide expression. For example, to treat
CC   disorders associated with decreased expression by rectifying mutations
CC   or deletions in a patient's genome that affect the activity of the
CC   polypeptides by expressing inactive proteins or to supplement the
CC   patient's own production of the polypeptide. Additionally, the nucleic
CC   acids may be used to produce the polypeptides, by inserting the nucleic
CC   acids into a host cell and culturing the cell to express the protein.
CC   The nucleic acid and its complementary sequences may also be used as DNA
CC   probes in diagnostic assays to detect and quantitate the presence of
CC   similar nucleic acid sequences in samples, and therefore which patients
CC   may be in need of restorative therapy. The proteins may also be used as
CC   antigens in the production of antibodies against bone marrow proteins
CC   and in assays to identify modulators of their expression and activity.
CC   The anti-bone marrow protein antibodies and antagonists may also be used
CC   to down regulate expression and activity. The antibodies may also be used
CC   as diagnostic agents for detecting the presence of the protein in samples
CC   (e.g. by enzyme linked immunosorbant assay (ELISA)). The proteins
CC   may be used to regulate haematopoiesis activity, and consequently in the
CC   treatment of myeloid or lymph cell disorders; in tissue regeneration,
CC   such as wound healing; as a nutritional supplement; and in treatment of
CC   immune disorders such as severe combined immunodeficiency (SCID).
XX
SQ   Sequence   43 AA,

  Query Match           100.0%;  Score 6;  DB 22;  Length 43;
  Best Local Similarity   66.7%;  Pred. No. 29;
  Matches    4;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 WXXWHF 6
           |::|||
Db      14 WLIWHF 19


RESULT 10
ABG64892
ID   ABG64892 standard; Protein; 53 AA.
XX
AC   ABG64892;
XX
DT   27-AUG-2002  (first entry)
XX
DE   Human albumin fusion protein #1567.
XX
KW   Albumin fusion protein; therapeutic protein X; human albumin; HA;
KW   human serum albumin; HSA; cancer; reproductive disorder;
KW   digestive disorder; immune disorder; endocrine disorder;
KW   haematopoietic disorder; neural disorder; connective disorder;
KW   cytostatic; antiinfertility; antiinflammatory; antiulcer;
KW   immunomodulator; anti-HIV; antidiabetic; haemostatic; nootropic;
KW   neuroprotective; antiparkinsonian; antimicrobial; neuroleptic;
KW   osteopathic; antiarthritic.
XX
OS   Homo sapiens.
OS   Synthetic.
XX
PN   WO200177137-A1.
XX
PD   18-OCT-2001.
XX
PF   12-APR-2001; 2001WO-US11988.
XX
PR   12-APR-2000; 2000US-229358P.
PR   25-APR-2000; 2000US-199384P.
PR   21-DEC-2000; 2000US-256931P.
XX
PA   (HUMA ) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Haseltine WA;
XX
DR   WPI; 2002-010886/01
XX
PT   New fusion protein for treating disease e.g. diabetes comprises an
PT   albumin fused to a therapeutic protein -
XX
PS   Claim 1; Page 1582; 2102pp; English.
XX
CC   The present invention relates to albumin fusion proteins comprising a
CC   therapeutic protein X and human albumin (HA, also known as human serum
CC   albumin, HSA). The proteins are useful for treating a disease or
CC   disorder that may be modulated by therapeutic protein X. The albumin
CC   extends the shelf life of protein X, and may increase its biological
CC   in vitro/in vivo activity. The protein is useful for treating and
CC   diagnosing disorders such as cancer, reproductive disorders, digestive
CC   disorders (e.g. Crohn's disease, ulcerative colitis), immune disorders
CC   (e.g. acquired immunodeficiency syndrome, AIDS), endocrine disorders
CC   (e.g. diabetes), haematopoietic disorders, neural disorders
CC   (e.g. Alzheimer's, Parkinson's, Creutzfeldt-Jacob disease,
CC   encephalomyelitis, meningitis, schizophrenia), and connective disorders
CC   (e.g. osteoporosis, arthritis). ABG63326-ABG65518 represent albumin
CC   fusion proteins of the invention.
XX
SQ   Sequence   53 AA;

  Query Match           100.0%;  Score 6;  DB 23;  Length 53;
  Best Local Similarity   66.7%;  Pred. No. 35;
  Matches    4;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 WXXWHF 6
           |..|||
Db      21 WASWHF 26


RESULT 11
AAE21233
ID   AAE21233 standard; Protein; 53 AA.
XX
AC   AAE21233;
XX
DT   01-JUL-2002  (first entry)
XX
DE   Human gene 18 encoded secreted protein HNNBM45, SEQ ID NO:98.
```

```
XX
KW   Human, secreted protein; immune disorder; antiallergic; antirheumatic;
KW   rheumatoid arthritis; breast neoplasia; breast cancer; antiarthritic;
KW   neurological disease; Alzheimer's disease; Parkinson's disease; trauma;
KW   Tourette syndrome; encephalitis; cytostatic; haemostatic; anaemia; mania;
KW   antiinflammatory; ophthalmological; dermatological; immunostimulatory;
KW   immunomodulatory; immunosuppressive; antibacterial; antipsoriatic;
KW   gene therapy; autoimmune disease; Huntington's disease; meningitis;
KW   demyelinating disease; peripheral neuropathy; congenital malformation;
KW   spinal cord injury; peripheral neuropathy; ischaemia; perception;
KW   multiple sclerosis; infarction; haemorrhage; schizophrenia; dementia;
KW   depression; panic disorder; learning disability; ALS; feeding disorder;
KW   hyperproliferative disorder; sleep pattern; cardiovascular disorder;
KW   reproductive disorder; digestive system disorder; behavioural disorder.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..31
FT                   /label= Signal_peptide
FT   Protein         32..53
FT                   /label= Mature_secreted_protein
FT   Misc_difference 50
FT                   /label= Unknown
FT                   /note= "Encoded by NNA"
XX
PN   WO200216390-A1.
XX
PD   28-FEB-2002.
XX
PF   17-JAN-2001; 2001WO-US01435.
XX
PR   18-AUG-2000; 2000US-226182P.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA, Ruben SM, Baker KP, Birse CE, Lippert BA, Olsen HS,
PI   Moore PA, Wei P, Ebner R, Duan DP, Shi Y, Choi GH, Piccella M;
PI   Ni J;
XX
DR   WPI; 2002-304113/34.
DR   N-PSDB; AAD33734.
XX
PT   An isolated nucleic acid molecule (I) comprising a polynucleotide which
PT   encodes a polypeptide useful in the diagnosis and treatment of
PT   disorders e.g. immune disorders -
XX
PS   Claim 11; Page 504; 534pp; English.
XX
CC   AAD33692-AAD33736 represent cDNAs corresponding to 21 human secreted
CC   protein genes, and AAE21191-AAE21235 represent the proteins they encode.
CC   AAE21236-AAE21241 represent human secreted protein fragments. The genes
CC   and their corresponding secreted proteins are useful for preventing,
CC   treating or ameliorating medical conditions, e.g., by protein or gene
CC   therapy. Pathological conditions can be diagnosed by determining the
CC   amount of the new protein in a sample or by determining the presence of
CC   mutations in the new genes. Specific uses are described for each of the
CC   21 genes, based on the tissues in which they are most highly expressed,
CC   and include developing prognostic tests for the diagnosis or treatment of
CC   immune or autoimmune diseases e.g. AIDS (acquired immune deficiency
CC   syndrome), asthma, anaemia and rheumatoid arthritis, breast neoplasia
CC   and breast cancer, neurological diseases e.g. Alzheimer's disease,
CC   Parkinson's disease, Huntington's disease, Tourette syndrome,
CC   meningitis, demyelinating diseases, peripheral neuropathies, neoplasia,
CC   trauma, congenital malformations, spinal cord injuries, [illegible]
CC   neuropathies induced by neurotoxins, peripheral neuropathies, multiple
CC   sclerosis, ischaemia and infarction, haemorrhages, schizophrenia, mania,
CC   dementia, depression, panic disorder, learning disabilities, ALS,
CC   altered behaviours e.g. disorders in feeding, sleep patterns, balance
CC   and perception, encephalitis, disorders in cell proliferation,
CC   recovery, reproductive and digestive systems, behavioural disorders and
CC   hyperproliferative disorders. The present sequence represents a human
CC   secreted protein of the invention.
XX
SQ   Sequence   53 AA;

  Query Match             100.0%;  Score 6;  DB 23;  Length 53;
  Best Local Similarity   66.7%;  Pred. No. 35;
  Matches    4;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 WXXWHF 6
           |:||||
Db      21 WASWHF 26


RESULT 12
AAU66178
ID   AAU66178 standard; Protein; 57 AA.
XX
AC   AAU66178;
XX
DT   27-FEB-2002  (first entry)
XX
DE   Propionibacterium acnes immunogenic protein #27074.
XX
KW   SAPHO syndrome; synovitis; acne; pustulosis; hyperostosis; osteomyelitis;
KW   uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
KW   inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
KW   dermatological; osteopathic; neuroprotectant.
XX
OS   Propionibacterium acnes.
XX
PN   WO200181581-A2.
XX
PD   01-NOV-2001.
XX
PF   20-APR-2001; 2001WO-US12865.
XX
PR   21-APR-2000; 2000US-199047P.
PR   [illegible]-2000; 2000US-226841P.
PR   07-JUL-2000; 2000US-216747P.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Skeiky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;
PI   L'maisonneuve J, Zhang Y, Jen S, Carter D;
XX
DR   WPI; 2001-616774/71.
DR   N-PSDB; AAS59718.
XX
PT   Propionibacterium acnes polypeptides and nucleic acids useful for
PT   vaccinating against and diagnosing infections, especially useful for
PT   treating acne vulgaris
XX
PS   Example 1; SEQ ID NO 27074; 1168pp; English.
XX
CC   Polypeptides AAY82105-AAY82217 represent Propionibacterium acnes immunogenic
CC   polypeptides. The proteins and their associated DNA sequences are useful in
CC   the treatment, prevention and diagnosis of medical conditions caused by
CC   P. acnes. The disorders include SAPHO syndrome (synovitis, acne,
CC   pustulosis, hyperostosis and osteomyelitis), uveitis and endophthalmitis.
CC   P. acnes is involved in infection of bone, joints and the central
CC   nervous system, however it is particularly involved in the inflammatory
CC   lesions associated with acne vulgaris. A method for detecting the
CC   presence or absence of P. acnes in a patient comprises contacting a
CC   sample with a binding agent that binds to the proteins of the invention
CC   and determining the amount of the polypeptide in the sample. The
CC   polypeptides may be used to antigens in the preparation of antibodies
CC   specific for P. acnes proteins. These antibodies can be used to
CC   downregulate expression and activity of P. acnes polypeptides and
CC   therefore treat P. acnes infections. The antibodies may also be used as
CC   diagnostic agents for determining P. acnes presence, for example, by
CC   enzyme linked immunosorbent assay (ELISA).
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
```

```
XX
SQ   Sequence   57 AA;

  Query Match           100.0%;  Score 6;  DB 22;  Length 57;
  Best Local Similarity   66.7%;  Pred. No. 37;
  Matches    4;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 WXXWHF 6
          |::|||
Db      4 WPPWHF 9


RESULT 13
ABP07322
ID   ABP07322 standard; Protein; 65 AA.
XX
AC   ABP07322;
XX
DT   24 JUN 2002  (first entry)
XX
DE   Human ORFX protein sequence SEQ ID NO:14626.
XX
KW   Human; open reading frame; ORFX; gene therapy; cancer; cirrhosis;
KW   hyperproliferative disorder; psoriasis; benign tumour, haemorrhage,
KW   degenerative disorder; osteoarthritis; neurodegenerative disorder;
KW   cardiovascular disease, diabetes mellitus, systemic lupus erythematosus,
KW   hypertension; hypothyroidism; cholesterol ester storage disease;
KW   immune deficiency; immune disorder; infectious disease;
KW   autoimmune disorder, rheumatoid arthritis, autoimmune thyroiditis,
KW   myasthenia gravis
XX
OS   Homo sapiens
XX
PN   WO200192523-A2.
XX
PD   06-DEC-2001.
XX
PF   29-MAY-2001; 2001WO-US10836.
XX
PR   30-MAY-2000; 2000US-206132P.
PR   29-AUG-2000; 2000US-228716P.
XX
PA   (CURA-) CURAGEN CORP
XX
PI   Shimkets RA,  Leach MD;
XX
DR   WPI; 2002-106308/14.
DR   N-PSDB; ABN23074.
XX
PT   Novel human polypeptides and polynucleotides useful for diagnosing,
PT   preventing and treating cardiovascular disease, neurodegenerative,
PT   hyperproliferative disorders and autoimmune disorders
XX
PS   Disclosure; SEQ ID 14626; 1037pp; English.
XX
CC   The present invention describes substantially purified human proteins
CC   (referred to as open reading frame, ORFX, where X is 1-11491 (see Table 1
CC   in the specification). APN19762 to ARN27252 encode the human ORFX
CC   proteins given in ABP00010 to ABP11500. ORFX proteins are useful for
CC   treating or preventing a pathology associated with an ORFX-associated
CC   disorder in humans, and in the manufacture of a medicament for treating a
CC   syndrome associated with ORFX-associated disorder. ORFX polynucleotide
CC   sequences can be used in gene therapy. ORFX sequences can be used in the
CC   treatment of cancer, hyperproliferative disorders, cirrhosis of liver,
CC   psoriasis, benign tumours, keloid, degenerative disorders, haemorrhage,
CC   osteoarthritis, neurodegenerative disorders, disorders related to organ
CC   transplantation, cardiovascular diseases, diabetes mellitus, systemic
CC   lupus erythematosus, hypertension, hypothyroidism, cholesterol ester
CC   storage disease, various immune deficiencies and disorders, infectious
CC   diseases, autoimmune disorders such as multiple sclerosis, rheumatoid
CC   arthritis, autoimmune thyroiditis, myasthenia gravis, graft-versus-host
CC   disease and autoimmune inflammatory eye disease. ORFX proteins are also
CC   useful for treating burns, incisions, ulcers, for treating osteoporosis,
CC   bone degenerative disorders, or periodontal disease, and for gut
CC   protection or regeneration and treatment of lung or liver fibrosis,
CC   reperfusion injury in various tissues and conditions resulting from
CC   systemic cytokine damage.
CC   N.B.  The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   65 AA;

  Query Match           100.0%;  Score 6;  DB 23;  Length 65;
  Best Local Similarity   66.7%;  Pred. No. 41;
  Matches    4;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 WXXWHF 6
          |::|||
Db     27 WKSWHF 32


RESULT 14
AAU56391
ID   AAU56391 standard; Protein; 87 AA.
XX
AC   AAU56391;
XX
DT   27-FEB-2002  (first entry)
XX
DE   Propionibacterium acnes immunogenic protein #17287.
XX
KW   SAPHO syndrome; synovitis; acne; pustulosis; hyperostosis; osteomyelitis;
KW   uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
KW   inflammatory lesion, acne vulgaris, enzyme linked immunosorbent assay;
KW   dermatological; osteopathic; neuroprotectant.
XX
OS   Propionibacterium acnes.
XX
PN   WO200181581-A2
XX
PD   01-NOV-2001.
XX
PF   20-APR-2001; 2001WO-US12865.
XX
PR   21 APR-2000; 2000US-198047P.
PR   02 JUN 2000; 2000US-208841P.
PR   07 JUL 2000; 2000US-216747P.
XX
PA   (CORI ) CORIXA CORP.
XX
PI   Skeiky YAW,  Persing DH,  Mitcham JL,  Wang SS,  Bhatia A;
PI   L'maisonneuve J,  Zhang Y,  Sun J,  Carter D;
XX
DR   WPI; 2001-616774/71.
DR   N-PSDB; AAS59575.
XX
PT   Propionibacterium acnes polypeptides and nucleic acids useful for
PT   vaccinating against and diagnosing infections, especially useful for
PT   treating acne vulgaris -
XX
PS   Example 1; SEQ ID No 17586; 1069pp; English.
XX
CC   Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic
CC   polypeptides. The proteins and their associated DNA sequences are used in
CC   the treatment, prevention and diagnosis of medical conditions caused by
CC   P. acnes. The disorders include SAPHO syndrome (synovitis, acne,
CC   pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.
CC   P. acnes is also involved in infections of bone, joints and the central
CC   nervous system, however it is particularly involved in the inflammatory
CC   lesions associated with acne vulgaris. A method for detecting the
CC   presence or absence of P. acnes in a patient comprises contacting a
CC   sample with a binding agent that binds to the proteins of the invention
CC   and determining the amount of bound protein in the sample. The
CC   polypeptides may be used as antigens in the production of antibodies
CC   specific for P. acnes proteins. These antibodies can be used to
```

CC downregulate expression and activity of P. acnes polypeptides and
CC therefore treat P. acnes infections. The antibodies may also be used as
CC diagnostic agents for determining P. acnes presence, for example, by
CC enzyme linked immunosorbent assay (ELISA).
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 97 AA;

Query Match 100.0%; Score 6; DB 22; Length 97;
Best Local Similarity 66.7%; Pred. No. 53;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 WXXWHF 6
:::|||
Db 34 WFFWHF 39

RESULT 15
AAU14777
ID AAU14777 standard; Protein; 96 AA.
XX
AC AAU14777;
XX
DT 24-OCT-2001 (first entry)
XX
DE Novel bone marrow polypeptide #176.
XX
KW Bone marrow; diagnostic; therapeutic; gene therapy; antigenic;
KW haematopoiesis; myeloid; lymph cell disorder; tissue regeneration;
KW wound healing; nutritional supplement; immune disorder;
KW severe combined immunodeficiency; SCID.
XX
OS Homo sapiens.
XX
PN WO200157187-A2.
XX
PD 09-AUG-2001.
XX
PF 05-FEB-2001; 2001WO-US03782.
XX
PR 03-FEB-2000; 2000US-0496914.
PR 20-JUN-2000; 2000US-0598075.
PR 19-JUL-2000; 2000US-0620325.
PR 30-NOV-2000; 2000US-0250683.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Ford JE, Doyle BC, Tang YT, Liu C, Anandi V, Zhou P, Xue AJ;
PI Ren F, Drmanac RT;
XX
DR WPI; 2001-489875/53.
DR N-PSDB; AAS23082.
XX
PT Nucleic acids encoding bone marrow polypeptides, useful in diagnostic
PT and gene therapy -
XX
PS Claim 10; Page 137; 392pp; English.
XX
CC AAU14602 AAU14794 represent novel bone marrow polypeptides of the
CC invention. The proteins and corresponding coding sequences may be used
CC in the prevention, diagnosis and treatment of diseases associated with
CC inappropriate bone marrow polypeptide expression. For example, to treat
CC disorders associated with decreased expression by rectifying mutations
CC or deletions in a patient's genome that affect the activity of the
CC polypeptides by expressing inactive proteins or to supplement the
CC patient's own production of the polypeptide. Additionally, the nucleic
CC acids may be used to produce the polypeptides, by inserting the nucleic
CC acids into a host cell and culturing the cell to express the protein.
CC The nucleic acid and its complementary sequence may also be used as DNA
CC probes in diagnostic assays to detect and quantitate the presence of
CC similar nucleic acid sequences in samples, and therefore which patients
CC may be in need of restorative therapy. The proteins may also be used as
CC antigens in the production of antibodies against bone marrow proteins
CC and in assays to identify modulators of their expression and activity.
CC The anti bone marrow protein antibodies and antagonists may also be used
CC to down regulate expression and activity. The antibodies may also be used
CC as diagnostic agents for detecting the presence of the protein in samples
CC (e.g. by enzyme linked immunosorbent assay (ELISA)). The proteins
CC may be used to regulate haematopoietic activity, and consequently in the
CC treatment of myeloid or lymph cell disorders, in tissue regeneration,
CC such as wound healing; as a nutritional supplement; and in treatment of
CC immune disorders such as severe combined immunodeficiency (SCID).
XX
SQ Sequence 96 AA;

Query Match 100.0%; Score 6; DB 22; Length 96;
Best Local Similarity 66.7%; Pred. No. 57;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 WXXWHF 6
:::|||
Db 67 WLIWHF 72

Search completed: February 20, 2003, 11:26:32
Job time : 29.3333 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 20, 2003, 19:35:48 ; Search time 24.3333 Seconds
(without alignments)
27.256 Million cell updates/sec

Title: US-09-912-414-9
Perfect score: 31
Sequence: 1 WXXWXF 6

Scoring table: BLOSUM62DX
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: A_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	31	100.0	6	15	AAR57386	Peptide for treati
2	31	100.0	6	15	AAR57391	Peptide for treati
3	31	100.0	6	21	AAB01492	Peptide which bind
4	31	100.0	6	21	AAB01497	Peptide which bind
5	31	100.0	6	21	AAB01499	Peptide which bind
6	31	100.0	6	21	AAB01505	Peptide which bind
7	31	100.0	6	21	AAB01506	Peptide which bind
8	31	100.0	6	21	AAB01508	Peptide which bind
9	31	100.0	7	22	AAE12181	Polyglutamine repe
10	31	100.0	7	22	AAM45777	H11 binding site c
11	31	100.0	8	15	AAR57399	Peptide for treati
12	31	100.0	9	21	AAB01498	Peptide which bind
13	31	100.0	9	22	AAU43677	Translocyte colony
14	31	100.0	10	15	AAR57400	Peptide for treati
15	31	100.0	11	19	AAW62149	Haemophilus influe
16	31	100.0	11	21	AAB20714	Polymeric immunogl
17	31	100.0	12	15	AAR57401	Peptide for treati
18	31	100.0	12	21	AAY88154	Oregon green 514 b
19	31	100.0	12	21	AAY88154	Oregon green 514 b
20	31	100.0	13	18	AAW38112	Dystrophin WW doma
21	31	100.0	13	23	ABP46201	Human BLyS binding
22	31	100.0	14	22	AAE07760	Human HLA DP restr
23	31	100.0	15	20	AAY30351	Epitope derived fr
24	31	100.0	15	21	AAY65508	Oestrogen receptor
25	31	100.0	15	23	AAU77026	Human ATP-dependen
26	31	100.0	15	23	AAU86240	Oestrogen receptor
27	31	100.0	15	23	AAE13229	Streptococcus pneu
28	31	100.0	17	21	AAB00193	Hepatitis GB virus
29	31	100.0	19	22	AAB76504	Peptide which bind
30	31	100.0	20	23	AAU90618	Insulin/insulin li
31	31	100.0	20	23	AAU30343	Insulin/insulin li
32	31	100.0	20	23	AAU70773	Hepatitis B virus
33	31	100.0	21	22	ABG04111	Novel human diagno
34	31	100.0	21	22	ABG29607	Novel human diagno
35	31	100.0	22	22	ABB78000	Consensus signal p
36	31	100.0	22	22	AAE13402	Consensus signal p
37	31	100.0	22	22	AAM52570	Consensus signal p
38	31	100.0	22	22	AAE13132	Consensus signal p
39	31	100.0	22	22	AAE12406	Consensus signal p
40	31	100.0	22	23	AAU99514	Consensus signal p
41	31	100.0	22	23	ABG63324	Synthetic signal p
42	31	100.0	22	23	AAE20851	Consensus signal p
43	31	100.0	22	23	AAU79154	Synthetic consensu
44	31	100.0	22	23	AAU75223	Consensus signal p
45	31	100.0	22	23	AAU75412	Tumour necrosis fa

ALIGNMENTS

RESULT 1
AAR57386
ID AAR57386 standard; peptide; 6 AA.
XX
AC AAR57386;
XX
DT 21-MAR-1995 (first entry)
XX
DE Peptide for treating diseases related to anti-DNA antibodies.
XX
KW Carrier, absorbing agent, treatment, anti-DNA antibody, immune complex.
XX
OS Synthetic.
XX
PN JP06192290-A.
XX
PD 12-JUL-1994.
XX
PF 18-JAN-1993; 93JP-0006098.
XX
PR 30-SEP-1992; 92JP-0261821.
XX
PA (KURS) KURARAY CO LTD.
XX
DR WPI; 1994-260510/32.
XX
PT A peptide and an adsorbing agent prepd. by immobilising it on a
PT carrier - useful for treatment of diseases related to anti-DNA
PT antibodies and immune complexes
XX
PS Disclosure; Page 11; 14pp; Japanese.
XX
CC The sequences given in AAR57386-413 are peptides which are all covered

CC by the claimed generic formula:
CC H-X-(A-B)n-Y-Z
CC A = Trp, Phe or a peptide fragment consisting of 2 residues;
CC B = Trp, Phe, Asn or Glu;
CC X and Y = a bond or Asp, Glu, Arg, Lys, His or a peptide fragment
CC consisting of 2-10 residues, provided that at least one of
CC X or Y are present;
CC Z = OH or NH2; and
CC n = 2-5.
CC These peptides may be immobilised on a carrier in the preparation of an
CC absorbing agent which may be used in the treatment of diseases related
CC to anti-DNA antibodies and/or immune complex.
XX
SQ Sequence 6 AA;

Query Match 100.0%; Score 31; DB 15; Length 6;
Best Local Similarity 50.0%; Pred. No. 7.8e+05;
Matches 3; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 WXXWXF 6
|::|:|
Db 1 WFFWFF 6

RESULT 2
AAR57391
ID AAR57391 standard; Protein; 6 AA.
XX
AC AAR57391;
XX
DT 21-MAR 1995 (first entry)
XX
DE Peptide for treating diseases related to anti-DNA antibodies.
XX
KW Carrier, absorbing agent, treatment, anti DNA antibody, immune complex.
XX
OS Synthetic.
XX
PN JP06192290-A
XX
PD 12-JUL-1994.
XX
PF 18-JAN 1993; 93JP-0006036.
XX
PR 30-SEP-1992; 92JP-0261821.
XX
PA (KURS) KURARAY CO LTD
XX
DR WPI; 1994-260810/32.
XX
PT A peptide and an absorbing agent prepd. by immobilising it on a
PT carrier useful for treatment of diseases related to anti-DNA
PT antibodies and immune complexes
XX
PS Disclosure; Page 11; 14pp; Japanese.
XX
CC The sequences given in AAR57390-413 are peptides which are all covered
CC by the claimed generic formula:
CC H-X-(A-B)n-Y-Z
CC A = Trp, Phe or a peptide fragment consisting of 2 residues;
CC B = Trp, Phe, Asn or Glu;
CC X and Y = a bond or Asp, Glu, Arg, Lys, His or a peptide fragment
CC consisting of 2-10 residues, provided that at least one of
CC X or Y are present;
CC Z = OH or NH2; and
CC n = 2-5.
CC These peptides may be immobilised on a carrier in the preparation of an
CC absorbing agent which may be used in the treatment of diseases realated
CC to anti-DNA antibodies and/or immune complex.
XX
SQ Sequence 6 AA;

Query Match 100.0%; Score 31; DB 15; Length 6;
Best Local Similarity 50.0%; Pred. No. 7.8e+05;
Matches 3; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 WXXWXF 6
|::|:|
Db 1 WFFWFF 6

RESULT 3
AAB01492
ID AAB01492 standard; peptide; 6 AA.
XX
AC AAB01492;
XX
DT 08-NOV-2000 (first entry)
XX
DE Peptide which binds to transcription factor E2F 1 DNA binding domain.
XX
KW DNA binding, transcription factor, E2F, E2F 1; cell cycle; DP-1;
KW activation, transcription, apoptosis; proliferative disorder;
KW psoriasis; restenosis.
XX
OS Synthetic.
XX
PN WO200044771-A1.
XX
PD 03-AUG-2000.
XX
PF 26-JAN 2000; 2000WO-GB00227.
XX
PR 28 JAN 1999; 99GB-0001810.
XX
PA (FROL) PROLIFIX LTD.
XX
PI Mueller R, Furtermann RE, Montigland G;
XX
DR WPI; 2000-532806/48.
XX
PT Peptide binding to the DNA binding domain of transcription factor E2F
PT and inhibiting cell cycle progression, useful for the treatment of
PT cancer
XX
PS Claim 6; Page 2; 42pp; English.
XX
CC Peptides which bind to the DNA binding domain of transcription
CC factor E2F and inhibit cell cycle progression may be useful as
CC research agents to investigate the interaction between E2F and DP-1,
CC or the activation of transcription by E2F 1/DP 1 heterodimers. They
CC may also be used for inducing apoptosis and/or cell cycle arrest in
CC a cell, particularly for treatment of cancer or other proliferative
CC disorders such as psoriasis and restenosis.
XX
SQ Sequence 6 AA;

Query Match 100.0%; Score 31; DB 31; Length 6;
Best Local Similarity 50.0%; Pred. No. 7.8e+05;
Matches 3; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 WXXWXF 6
|::|:|
Db 1 WVRWHF 6

RESULT 4
AAB01497
ID AAB01497 standard; peptide; 6 AA.
XX
AC AAB01497;
XX
DT 08-NOV-2000 (first entry)
XX
DE Peptide which binds to transcription factor E2F 1 DNA binding domain.
XX

```
KW   DNA binding; transcription factor, E2F, E2F 1, cell cycle, DP-1,
KW   activation; transcription, apoptosis, proliferative disorder,
KW   psoriasis; restenosis.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 2
FT                   /note= "Any amino acid"
FT   Misc-difference 3
FT                   /note= "Any amino acid"
FT   Misc-difference 5
FT                   /note= "Any amino acid"
XX
PN   WO200044771-A1.
XX
PD   03-AUG-2000
XX
PF   26-JAN-2000; 2000WO-GB00227.
XX
PR   26-JAN-1999;   99GB-0001710.
XX
PA   (PROL-) PROLIFIX LTD.
XX
PI   Mueller R,  Kontermann RE,  Montigiani S;
XX
DR   WPI; 2000-532806/48.
XX
PT   Peptides binding to the DNA binding domain of transcription factor E2F
PT   and inhibiting cell cycle progression, useful for the treatment of
PT   cancer
XX
PS   Claim 4; Page 9; 42pp; English.
XX
CC   Peptides which bind to the DNA binding domain of transcription
CC   factor E2F and inhibit cell cycle progression may be useful as
CC   research agents to investigate the interaction between E2F and DP 1,
CC   or the activation of transcription by E2F-1/DP-1 heterodimers. They
CC   may also be used for inducing apoptosis and/or cell cycle arrest in
CC   a cell, particularly for treatment of cancer or other proliferative
CC   disorders such as psoriasis and restenosis.
XX
SQ   Sequence   6 AA;

  Query Match           100.0%;  Score 31;  DB 21;  Length 6;
  Best Local Similarity  100.0%;  Pred. No. 7.8e+05;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 WXXWXF 6
          ||||||
Db      1 WXXWXF 6


RESULT 5
AAB01499
ID   AAB01499 standard; peptide; 6 AA.
XX
AC   AAB01499;
XX
DT   08-NOV 2000  (first entry)
XX
DE   Peptide which binds to transcription factor E2F-1 DNA binding domain
XX
KW   DNA binding, transcription factor, E2F, E2F 1, cell cycle, DP 1,
KW   activation; transcription; apoptosis; proliferative disorder;
KW   psoriasis; restenosis.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 2
FT                   /note= "Any amino acid"
FT   Misc-difference 3
FT                   /note= "Any amino acid"
XX
PN   WO200044771-A1.
XX
PD   03-AUG-2000
XX
PF   26-JAN-2000; 2000WO-GB00227.
XX
PR   26-JAN-1999;   99GB-0001710.
XX
PA   (PROL-) PROLIFIX LTD.
XX
PI   Mueller R,  Kontermann RE,  Montigiani S;
XX
DR   WPI; 2000-532806/48.
XX
PT   Peptides binding to the DNA binding domain of transcription factor E2F
PT   and inhibiting cell cycle progression, useful for the treatment of
PT   cancer
XX
PS   Claim 4; Page 9; 42pp; English.
XX
CC   Peptides which bind to the DNA binding domain of transcription
CC   factor E2F and inhibit cell cycle progression may be useful as
CC   research agents to investigate the interaction between E2F and DP 1,
CC   or the activation of transcription by E2F-1/DP-1 heterodimers. They
CC   may also be used for inducing apoptosis and/or cell cycle arrest in
CC   a cell, particularly for treatment of cancer or other proliferative
CC   disorders such as psoriasis and restenosis.
XX
SQ   Sequence   6 AA;

  Query Match           100.0%;  Score 31;  DB 21;  Length 6;
  Best Local Similarity  83.3%;  Pred. No. 7.8e+05;
  Matches    5;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 WXXWXF 6
          ||||:|
Db      1 WXXWHF 6


RESULT 6
AAB01505
ID   AAB01505 standard; peptide; 6 AA.
XX
AC   AAB01505;
XX
DT   08-NOV-2000  (first entry)
XX
DE   Peptide which binds to transcription factor E2F-1 DNA binding domain.
XX
KW   DNA binding; transcription factor; E2F; E2F-1; cell cycle; DP-1;
KW   activation; transcription; apoptosis; proliferative disorder;
KW   psoriasis; restenosis.
XX
OS   Synthetic.
XX
PN   WO200044771-A1.
XX
PD   03-AUG-2000
XX
PF   26-JAN-2000; 2000WO-GB00227.
XX
PR   26-JAN-1999;   99GB-0001710.
XX
PA   (PROL-) PROLIFIX LTD.
XX
PI   Mueller R,  Kontermann RE,  Montigiani S;
XX
DR   WPI; 2000-532806/48.
XX
PT   Peptides binding to the DNA binding domain of transcription factor E2F
PT   and inhibiting cell cycle progression, useful for the treatment of
```

PT cancer
XX
PS Example, Page 26, 42pp, English.
XX
CC Peptides which bind to the DNA binding domain of transcription
CC factor E2F and inhibit cell cycle progression may be useful as
CC research agents to investigate the interaction between E2F and DP-1,
CC or the activation of transcription by E2F 1/DP 1 heterodimers. They
CC may also be used for inducing apoptosis and/or cell cycle arrest in
CC a cell, particularly for treatment of cancer or other proliferative
CC disorders such as psoriasis and restenosis.
XX
SQ Sequence 6 AA;

Query Match 100.0%; Score 31; DB 21; Length 6;
Best Local Similarity 83.3%; Pred. No. 7.8e+05;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 WXXWXF 6
|::|:|
Db 1 WARWHF 6

RESULT 7
AAB01506
ID AAB01506 standard; peptide; 6 AA.
XX
AC AAB01506;
XX
DT 08-NOV-2000 (first entry)
XX
DE Peptide which binds to transcription factor E2F-1 DNA binding domain.
XX
KW DNA binding, transcription factor, E2F, E2F-1, cell cycle, DP-1,
KW activation, transcription, apoptosis, proliferative disorder,
KW psoriasis, restenosis.
XX
OS Synthetic.
XX
PN WO200044771-A1.
XX
PD 03-AUG-2000.
XX
PF 26-JAN-2000; 2000WO-GB00227.
XX
PR 26-JAN-1999; 99GB-0001710.
XX
PA (PROL-) PROLIFIX LTD.
XX
PI Mueller R, Kontermann RE, Montigiani S;
XX
DR WPI; 2000-511256/48.
XX
PT Peptides binding to the DNA binding domain of transcription factor E2F
PT and inhibiting cell cycle progression, useful for the treatment of
PT cancer
XX
PS Example; Page 26; 42pp; English.
XX
CC Peptides which bind to the DNA binding domain of transcription
CC factor E2F and inhibit cell cycle progression may be useful as
CC research agents to investigate the interaction between E2F and DP-1,
CC or the activation of transcription by E2F 1/DP 1 heterodimers. They
CC may also be used for inducing apoptosis and/or cell cycle arrest in
CC a cell, particularly for treatment of cancer or other proliferative
CC disorders such as psoriasis and restenosis.
XX
SQ Sequence 6 AA;

Query Match 100.0%; Score 31; DB 21; Length 6;
Best Local Similarity 83.3%; Pred. No. 7.8e+05;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 WXXWXF 6
|::|:|
Db 1 WVAWHF 6

RESULT 8
AAB01508
ID AAB01508 standard; peptide; 6 AA.
XX
AC AAB01508;
XX
DT 08-NOV-2000 (first entry)
XX
DE Peptide which binds to transcription factor E2F-1 DNA binding domain.
XX
KW DNA binding; transcription factor, E2F, E2F-1, cell cycle; DP-1;
KW activation, transcription, apoptosis, proliferative disorder;
KW psoriasis; restenosis.
XX
OS Synthetic.
XX
PN WO200044771-A1.
XX
PD 03-AUG-2000.
XX
PF 26-JAN-2000; 2000WO-GB00227.
XX
PR 26-JAN-1999; 99GB-0001710.
XX
PA (PROL-) PROLIFIX LTD.
XX
PI Mueller R, Kontermann RE, Montigiani S;
XX
DR WPI; 2000-511256/48.
XX
PT Peptides binding to the DNA binding domain of transcription factor E2F
PT and inhibiting cell cycle progression, useful for the treatment of
PT cancer
XX
PS Example; Page 26; 42pp; English.
XX
CC Peptides which bind to the DNA binding domain of transcription
CC factor E2F and inhibit cell cycle progression may be useful as
CC research agents to investigate the interaction between E2F and DP-1,
CC or the activation of transcription by E2F 1/DP 1 heterodimers. They
CC may also be used for inducing apoptosis and/or cell cycle arrest in
CC a cell, particularly for treatment of cancer or other proliferative
CC disorders such as psoriasis and restenosis.
XX
SQ Sequence 6 AA;

Query Match 100.0%; Score 31; DB 21; Length 6;
Best Local Similarity 83.3%; Pred. No. 7.8e+05;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 WXXWXF 6
|::|:|
Db 1 WVRWAF 6

RESULT 9
AAB12181
ID AAB12181 standard; peptide; 7 AA.
XX
AC AAB12181;
XX
DT 18-DEC-2001 (first entry)
XX
DE Polyglutamine repeat domain binding peptide, QPB1-M7FD.
XX
KW Polyglutamine repeat domain, huntingtin, atrophin, ataxin;
KW androgen receptor protein, neurodegenerative disease, anticonvulsant;
KW Huntington's disease; dentatorubral pallidoluysian atrophy; DRPLA;

KW spinobulbar muscular atrophy; spinocerebellar ataxia; nootropic.
XX
OS Unidentified.
XX
PN WO200168678-A2.
XX
PD 20-SEP-2001.
XX
PF 14-MAR-2001; 2001WO-US08222.
XX
PR 16-MAR-2000; 2000US-189781P.
XX
PA (UYDU-) UNIV DUKE.
XX
PI Strittmatter WJ, Burke JR, Nagai Y;
XX
DR WPI; 2001-616245/71.
XX
PT New polypeptides which selectively bind to expanded polyglutamine
PT repeat domains, useful for treating neurodegenerative diseases, e.g.
PT Huntington's disease -
XX
PS Example 13; Page 14; 59pp; English
XX
CC The present sequence is a polyglutamine repeat domain binding peptide.
CC The peptide is useful for treating a cell that contains and expresses a
CC protein having an expanded polyglutamine region (e.g. huntingtin;
CC atrophin 1, ataxin 1, 2, 6 or 7, or androgen receptor protein), or
CC treating a neurodegenerative disease characterised by the presence of
CC expanded polyglutamine repeats, e.g. Huntington's disease, dentatorubral
CC pallidoluysian atrophy (DRPLA), spinobulbar muscular atrophy, and
CC spinocerebellar ataxia types 1, 2, 3, 6 and 7
XX
SQ Sequence 7 AA;

Query Match 100.0%; Score 31; DB 22; Length 7;
Best Local Similarity 50.0%; Pred. No. 7.8e+05;
Matches 3; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 WXXWXF 6
|::|:|
Db 1 WKWWPF 6

RESULT 10
AAM45777
ID AAM45777 standard; Peptide; 7 AA.
XX
AC AAM45777;
XX
DT 26-OCT-2001 (first entry)
XX
DE H11 binding site consensus conforming peptide (CCP) #2048.
XX
KW Antigen-binding; tumour; diagnosis; stress protein-peptide complex; SPPC;
KW immunogenically cross-reactive; cancer; immunogenic cancer cell;
KW cytostatic; vaccine, tumour specific immunogenic response inducer;
KW astrocytoma, fibrosarcoma, myxosarcoma, liposarcoma, oligodendroglioma;
KW ependymoma; medulloblastoma; primitive neural ectodermal tumour.
XX
OS Homo sapiens
OS Synthetic.
XX
PN CA2290722-A1.
XX
PD 08-JUN-2001.
XX
PF 08-DEC-1999; 99CA-2290722
XX
PR 08-DEC-1999; 99CA-2290722.
XX
PA (NOVO) NOVOPHARM BIOTECH INC.
XX
PI Kaplan HA, Maiti PK, Fast DG, Herman W, Dan MD, Lewis KE;
PI Entwistle JM, MacDonald GC;
XX
DR WPI; 2001-425937/46.
XX
PT Composition useful for treating and diagnosing cancer, comprises stress
PT protein-peptide complexes associated with tumor, and isolated
PT antigen-binding fragments of an antibody that binds specifically to the
PT complex -
XX
PS Example 4; Page 108; 154pp; English.
XX
CC The present invention describes a composition (I) comprising stress
CC protein-peptide complexes (SPPC) associated with tumours that is
CC specifically immunogenically cross-reactive with cell surface associated
CC SPPCs specific to target cancer (TC). Also described is an isolated
CC antigen binding fragment of an antibody that binds specifically to SPPCs
CC or a population of different SPPCs consisting of immunogenic cancer cell
CC surface associated SPPC of TC. (I) has cytostatic activity and can be
CC used in vaccine production and as a tumour-specific immunogenic response
CC inducer. (I) is useful for treating 71 types of cancers or tumours in a
CC subject, such as astrocytoma, fibrosarcoma, myxosarcoma, liposarcoma,
CC oligodendroglioma, ependymoma, medulloblastoma, and primitive neural
CC ectodermal tumour (PNET). (I) is useful as cancer immunogen including
CC vaccines. (I) is useful for diagnostic and palliative use, for detecting
CC or imaging cancer cells, and to monitor the course of amelioration of
CC malignancy in an individual. AAM43707 to AAM47109 represent peptides
CC which are used in the exemplification of the present invention.
XX
SQ Sequence 7 AA;

Query Match 100.0%; Score 31; DB 22; Length 7;
Best Local Similarity 50.0%; Pred. No. 7.8e+05;
Matches 3; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 WXXWXF 6
|::|:|
Db 1 WRPWNF 6

RESULT 11
AAR57399
ID AAR57399 standard; Protein; 8 AA.
XX
AC AAR57399;
XX
DT 21-MAR-1995 (first entry)
XX
DE Peptide for treating diseases related to anti-DNA antibodies.
XX
KW Carrier; absorbing agent; treatment; anti-DNA antibody; immune complex.
XX
OS Synthetic.
XX
PN JP06192290-A.
XX
PD 12-JUL-1994.
XX
PF 18-JAN-1993; 93JP-0006098.
XX
PR 30-SEP-1992; 92JP-0261821.
XX
PA (KUPS) KURARAY CO LTD.
XX
DR WPI; 1994-260510/32.
XX
PT A peptide and an adsorbing agent prepd. by immobilising it on a
PT carrier useful for treatment of diseases related to anti DNA
PT antibodies and immune complexes
XX
PS Disclosure; Page 12; 14pp; Japanese.
XX
CC The sequences given in AAR57386-413 are peptides which are all covered

CC by the claimed generic formula.
CC H-X-(A-B)n-Y-Z
CC A = Trp, Phe or a peptide fragment consisting of 2 residues,
CC B = Trp, Phe, Asn or Gln,
CC X and Y = a bond or Asp, Glu, Arg, Lys, His or a peptide fragment
CC consisting of 2 10 residues, provided that at least one of
CC X or Y are present;
CC Z = OH or NH2, and
CC n = 2 5.
CC These peptides may be immobilised on a carrier in the preparation of an
CC absorbing agent which may be used in the treatment of diseases realated
CC to [illegible] antibodies and/or immune complex.
XX
SQ Sequence 8 AA;

Query Match 100.0%; Score 31; DB 15; Length 8;
Best Local Similarity 50.0%; Pred. No. 7.8e+05;
Matches 3; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 WXXWXF 6
|:::|:|
Db 1 WWPWWF 6

RESULT 12
AAB01498
ID AAB01498 standard; peptide; 9 AA.
XX
AC AAB01498;
XX
DT 08-NOV-2000 (first entry)
XX
DE Peptide which binds to transcription factor E2F-1 DNA binding domain.
XX
KW DNA binding, transcription factor, E2F, E2F-1, cell cycle, DP-1,
KW activation, transcription, apoptosis, proliferative disorder,
KW psoriasis; restenosis.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Misc-difference 2
FT /note= "Any amino acid"
FT Misc-difference 3
FT /note= "Any amino acid"
FT Misc-difference 5
FT /note= "Any amino acid"
FT Misc-difference 7
FT /note= "Any amino acid"
FT Misc-difference 8
FT /note= "Any amino acid"
XX
PN WO200044771-A1.
XX
PD 03 AUG-2000.
XX
PF 26-JAN-2000; 2000WO-GB00227.
XX
PR 26-JAN-1999; 99GB-0001716.
XX
PA (PROL-) PROLIFIX LTD.
XX
PI Mueller R, Kuntermann RE, Mcfarlane S;
XX
DR WPI; 2000-503806/48.
XX
PT Peptides binding to the DNA binding domain of transcription factor E2F
PT and inhibiting cell cycle progression, useful for the treatment of
PT cancer
XX
PS Claim 4; Page 9; 42pg; English.
XX
CC Peptides which bind to the DNA binding domain of transcription
CC factor E2F and inhibit cell cycle progression may be useful as
CC research agents to investigate the interaction between E2F and DP-1,
CC in the activation of transcription by E2F-1/DP-1 heterodimers. They
CC may also be used for inducing apoptosis and/or cell cycle arrest in
CC a cell, particularly for treatment of cancer or other proliferative
CC disorders such as psoriasis and restenosis.
XX
SQ Sequence 9 AA;

Query Match 100.0%; Score 31; DB 11; Length 9;
Best Local Similarity 100.0%; Pred. No. 7.8e+05;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 WXXWXF 6
||||||
Db 1 WXXWXF 6

RESULT 13
AAU93672
ID AAU93672 standard; Peptide; 9 AA.
XX
AC AAU93672;
XX
DT 02-JUL-2002 (first entry)
XX
DE Granulocyte colony stimulating factor receptor binding peptide #478.
XX
KW G-CSFR, granulocyte-colony stimulating factor receptor, cytokine;
KW haematopoietic growth factor; neutrophil proliferation; AIDS;
KW neutrophil differentiation; acquired immunodeficiency syndrome;
KW chemotherapy induced neutropaenia, community acquired pneumonia;
KW depressed neutrophil count, immunostimulant.
XX
OS Synthetic.
XX
PN WO2002057076 A2.
XX
PD 31 JAN-2002.
XX
PF 22-JAN-2002; 2002WO-US02046.
XX
PR 26 JUL 2001, 2001US-0308031.
XX
PA (SLAA) SLAAL THERAP LTD.
XX
PI Cwirla SE, Balu P, Duffin DJ, Piplani S, McBowen-Merrill B;
PI Schatz PJ;
XX
DR WPI; 2002-340848/36.
XX
PT Novel compounds, useful for treating depressed neutrophil count,
PT comprise peptide chains of approximately 6 to 40 amino acids in length
PT that bind to granulocyte colony stimulating factor receptor -
XX
PS Claim 41; Page 67; 90pp; English.
XX
CC The invention relates to compounds comprising a peptide chain
CC approximately 6 to 40 amino acids in length that binds to granulocyte-
CC colony stimulating factor receptor (G-CSFR). The compounds contain
CC specific sequences of the peptide appearing as AAU79402-AAU79406
CC and the generic sequences XV_1XV_2XV_3XV_4XV_5XV_6XV_7XV_8 (where
CC XV_1 = E, D, Q, V or Y; XV_2 = E, A, L, M, S, W or Q; XV_3 = K, R or T;
CC XV_4 = L, A or V, XV_5 = R, A, M, H, E, V, L, G, P, Q or S; XV_6 = E or
CC Y; XV_7 = A or S; and XV_8 = R, H, G or L; and XVI_1XVI_2XVI_3XVI_4XVI_5
CC EXVI_6XVI_7XVI_8XVI_9 (where XVI_1 = A, E or G; XVI_2 = F, H or D;
CC XVI_3 = F or S; XVI_4 = F, Y, M, N, L, E, C, T, S, I or F; XVI_5 = A, S or
CC T, XVI_6 = E, D, T, L, F or A, XVI_7 = R, W, H, L, G, A or Q; XVI_8 = R
CC or E; and XVI_9 = W, G or R). The compounds are used for treating
CC conditions associated with depressed neutrophil count e.g. chemotherapy
CC induced neutropaenia, AIDS induced neutropaenia or immunity acquired
CC pneumonia-induced pneumonia. The compounds are useful as in vitro as
CC tools for understanding the biological role of granulocyte-colony

```
CC   stimulating factor (G-CSF a haematopoietic growth factor and
CC   cytokine that stimulates neutrophil proliferation and differentiation),
CC   including evaluation of many factors thought to influence, and be
CC   influenced by, production of white blood cells, in the development of
CC   compounds that bind to G-CSFR, as reagents for detecting G-CSF receptor
CC   or related receptor on living cells, fixed cells, in biological fluid, in
CC   tissue homogenates or in purified natural biological materials, in situ
CC   staining, fluorescence-activated cell sorting (FACS), Western blotting or
CC   enzyme-linked immunoadsorptive assay (ELISA), in receptor purification or
CC   in purifying cells expressing G-CSFR on the cell surface (or inside
CC   permeabilised cells) as a commercial research reagent for various medical
CC   and diagnostic uses or to treat a disease that would benefit from the
CC   ability to of a compound to mimic the effects of G-CSF in vivo.
CC   The compounds bind specifically to G-CSFR and allow for studies of
CC   biological activities mediated by the receptor and for the treatment of
CC   diseases, disorders and conditions that would benefit from activating or
CC   inactivating G-CSFR. The present sequence is a G-CSFR binding peptide of
CC   the invention.
XX
SQ   Sequence   9 AA;

  Query Match             100.0%;  Score 31;  DB 23;  Length 9;
  Best Local Similarity   90.0%;  Pred. No. 7.8e-06;
  Matches    3;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 WXXWXF 6
          |::|:|
Db      4 WWGWNF 9


RESULT 14
AAR57400
ID   AAR57400 standard; Protein; 10 AA
XX
AC   AAR57400;
XX
DT   21-MAR-1995  (first entry)
XX
DE   Peptide for treating diseases related to anti-DNA antibodies
XX
KW   Carrier; absorbing agent; treatment; anti-DNA antibody; immune complex
XX
OS   Synthetic
XX
PN   JP06192290-A.
XX
PD   12-JUL-1994.
XX
PF   18-JAN-1993;   93JP-0006098.
XX
PR   30-SEP-1992;   92JP-0261921.
XX
PA   (KURS ) KURARAY CO LTD.
XX
DR   WPI; 1994-260510/32
XX
PT   A peptide and an adsorbing agent prepd. by immobilising it on a
PT   carrier - useful for treatment of diseases related to anti-DNA
PT   antibodies and immune complexes
XX
PS   Disclosure; Page 12; 14pp; Japanese
XX
CC   The sequences given in AAR57386-413 are peptides which are all covered
CC   by the claimed generic formula:
CC                          H-X-(A-B)n-Y-Z
CC   A = Trp, Phe or a peptide fragment consisting of 2 residues;
CC   B = Trp, Phe, Asn or Glu;
CC   X and Y = a bond or Asp, Glu, Arg, Lys, His or a peptide fragment
CC             consisting of 2-10 residues, provided that at least one of
CC             X or Y are present;
CC   Z = OH or NH2; and
CC   n = 2-5.
CC   These peptides may be immobilised on a carrier in the preparation of an
CC   absorbing agent which may be used in the treatment of diseases realated
CC   to anti-DNA antibodies and/or immune compex.
XX
SQ   Sequence   10 AA;

  Query Match             100.0%;  Score 31;  DB 15;  Length 10;
  Best Local Similarity   90.0%;  Pred. No. 1.9e-02;
  Matches    3;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 WXXWXF 6
          |::|:|
Db      3 WWFWWF 8


RESULT 15
AAW62149
ID   AAW62149 standard, peptide, 11 AA.
XX
AC   AAW62149;
XX
DT   16-SEP-1998  (first entry)
XX
DE   Haemophilus influenzae tyrosine tRNA synthetase binding peptide 22.
XX
KW   Identification; ligand; biological activity; target-binding;
KW   drug screening; library; inhibitory ligand.
XX
OS   Synthetic.
OS   Haemophilus influenzae.
XX
PN   WO9819162-A1.
XX
PD   07-MAY-1998.
XX
PF   31-OCT-1997;   97WO-US19638.
XX
PR   31-OCT-1996;   96US-0740671.
XX
PA   (NOVA-) NOVALON PHARM CORP.
XX
PT   Fowlkes DM, Frolinger JA, Hyde-Deruyscher RP, May RK;
XX
DR   WPI; 1998-272389/24.
XX
PT   Identifying ligands which mediate biological activity of a protein
PT   by identifying target-binding ligands and screening a library for
PT   ligands which inhibit target-binding ligand mediated activity
XX
PS   Example 5; Page 100; 143pp; English.
XX
CC   A method has been developed for identifying a ligand which mediates the
CC   biological activity of a target protein (T) by inhibiting the binding
CC   of (T) to a binding partner. The method comprises: (a) screening a first
CC   combinatorial library comprising first member ligands for binding to the
CC   target-binding ligands (TBLs), to identifying one or more TBLs; (b)
CC   screening a second library comprising second member ligands for the
CC   ability to inhibit the binding of one or more of the TBLs to the target
CC   protein, and so obtaining one or more inhibitory ligands; and (c)
CC   determining which of the inhibitory ligands can mediate a biological
CC   activity of the target protein. The present sequence represents a
CC   potential binding peptide for Haemophilus influenzae tyrosine tRNA
CC   synthetase from an example of the present invention. The method can be
CC   used for identifying drugs which can mediate the biological activity of
CC   a target protein. It can be used to identify the biological activity of
CC   a target protein whose biological function is not known and perhaps
CC   cannot be determined directly. The method can also be used to identify
CC   new inhibitory ligands of specific target proteins. The method provides
CC   high throughput screens which are essentially identical for similar and
CC   dissimilar targets, bypassing the need to develop distinct assays for
CC   biochemically diverse targets.
XX
SQ   Sequence   11 AA;
```

Query Match 100.0%; Score 31; DB 19; Length 11;
Best Local Similarity 50.0%; Pred. No. 2.1e+02;
Matches 3; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 WXXWXF 6
           |:::|
Db       5 WPDWGF 10
```

Search completed: February 20, 2003, 10:47:56
Job time : 31.3333 secs